

**Über eine abstandserzeugende Funktion
von Wahrscheinlichkeitsmaßen**

Inaugural-Dissertation
zur Erlangung des Grades eines
Doktors der Wirtschafts- und Sozialwissenschaften
bei der Wirtschafts- und Sozialwissenschaftlichen
Fakultät der Freien Universität Berlin

vorgelegt von
Diplom-Mathematiker Götz Trenkler
aus Dresden

1973

**Über eine abstandserzeugende Funktion
von Wahrscheinlichkeitsmaßen**

Inaugural-Dissertation
zur Erlangung des Grades eines
Doktors der Wirtschafts- und Sozialwissenschaften
bei der Wirtschafts- und Sozialwissenschaftlichen
Fakultät der Freien Universität Berlin

vorgelegt von
Diplom-Mathematiker Götz Trenkler
aus Dresden

Vorsitzender des Fachbereichs: N.N.

Referent: Prof. Dr. Max-Detlev Jöhnk

Korreferent: Prof. Dr. Hans Münzner

Tag der mündlichen Prüfung: 17. Dezember 1973



Inhaltsverzeichnis

Einleitung	Seite 2
I. Kapitel Grundlegende Eigenschaften	Seite 5
II. Kapitel Ein Hauptsatz	Seite 22
III. Kapitel Gestutzte Verteilungen	Seite 37
IV. Kapitel Suffizienz	Seite 48
V. Kapitel Produkträume	Seite 56
VI. Kapitel Grenzwertsätze	Seite 61
VII. Kapitel Anwendungen auf die Testtheorie	Seite 74
Anhang Einige wichtige AEF	Seite 98
Literaturverzeichnis	Seite 104



Digitized by the Internet Archive
in 2026

Über eine abstandserzeugende Funktion
von Wahrscheinlichkeitsmaßen

Einleitung

In einer Arbeit von Golomb [17] wird für ein diskretes Wahrscheinlichkeitsmaß P mit dem Träger $\mathbb{N}$ die reelle Funktion T des reellen Arguments u mit

$$T(u) = \sum_{n \in \mathbb{N}} [P(n)]^u \quad (0.1)$$

untersucht. Nach formaler Differentiation nach u stellt man fest, daß

$$-\left. \frac{dT(u)}{du} \right|_{u=1} = - \sum_{n \in \mathbb{N}} P(n) \log P(n) \quad (0.2)$$

gerade die bekannte Shannonsche Entropie liefert. Folgerichtig nennt Golomb $T(u)$ die "information generating function".

Dies gibt Anlaß, sich folgender Fragestellung zuzuwenden: Wenn zwei diskrete Wahrscheinlichkeitsverteilungen $P(x)$ und $Q(x)$ vorliegen, gibt es dann eine reelle Funktion einer reellen Variablen $I(t)$, für die

$$\left. \frac{d}{dt} I(t) \right|_{t=t_0} = K(P, Q) \quad (0.3)$$

für ein passendes $t_0 \in \mathbb{R}$, wobei

$$K(P, Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \quad (0.4)$$

das sogenannte Kullback-Leibler-Informationsmaß (Vgl. [40]) ist ?

Dieses Problem drängt sich auf, weil das Kullback-Leibler-Maß eine Verallgemeinerung der Shannonschen Entropie darstellt. Nach längerem Experimentieren leistete die Funktion

$$I(t) = \sum_x P^t(x) Q(x)^{1-t} \quad (0.5)$$

das Gewünschte. Es fällt auf, daß bei Kenntnis der Funktion $I(t)$ als "Nebenprodukt" der u.a. von Skarabis [52] in seiner Dissertation verwendete und von Bhattacharya [4] eingeführte Abstand zweier Wahrschein-

lichkeitsmaße P und Q :

$$I\left(\frac{1}{2}\right) = \int_X \sqrt{P(x) Q(x)} \quad (0.6)$$

mit abfällt.

Bekanntlich existiert in der einschlägigen Literatur eine Vielzahl von Abstandsmaßen zwischen Wahrscheinlichkeitsverteilungen. So verwenden Adhikari und Joshi [1] die folgenden Abstände:

$$\rho(\mu, \nu) = \int \sqrt{f_\mu(x) f_\nu(x)} d\lambda_0(x) \quad (0.7)$$

$$d_H(\mu, \nu) = \frac{1}{2} \int [\sqrt{f_\mu(x)} - \sqrt{f_\nu(x)}]^2 d\lambda_0(x) \quad (0.8)$$

$$d_J(\mu, \nu) = \left[\int (\sqrt{f_\mu(x)} - \sqrt{f_\nu(x)})^2 d\lambda_0(x) \right]^{\frac{1}{2}} \quad (0.9)$$

$$d_V(\mu, \nu) = \sup_{E \in \mathfrak{K}} |\mu(E) - \nu(E)| \quad (0.10)$$

wobei μ, ν Maße auf einer σ -Algebra $\mathfrak{K}$ sind, die durch das σ -endliche Maß λ_0 dominiert werden. f_μ bzw. f_ν sind die entsprechenden Radon-Nikodým-Dichten.

Rényi [48] und Csiszár [8] verwenden als Abstand:

$$I_\alpha(\mu || \nu) = \frac{1}{\alpha - 1} \log \int f_\mu(x)^\alpha f_\nu(x)^{1-\alpha} d\lambda_0(x) \quad (0.11)$$

mit $\alpha \neq 1$, $\alpha \in \mathbb{R}$.

$I_\alpha(\mu || \nu)$ heißt bei den Autoren der "Informationsgewinn α -ter Ordnung".

Zu nennen ist noch die Chernoffsche Divergenz [6]:

$$D(Y) = -\log \left[\inf_{0 < t < 1} \int f_\mu(x)^t f_\nu^{1-t} d\lambda_0(x) \right] \quad (0.12)$$

Im diskreten Fall hängen diese Abstände irgendwie von $I(t)$ ab. (Wie später zu sehen ist, auch im allgemeinen Fall !). Zur Aufhellung der Struktur des Verhältnisses zweier Wahrscheinlichkeitsmaße dürfte die Funktion $I(t)$ also von starkem Interesse sein. Das Studium der

Funktion $I(t)$ ist Gegenstand dieser Arbeit. Um die Untersuchungen so allgemein wie möglich zu halten, konnte auf maßtheoretische Hilfsmittel nicht verzichtet werden.

Im ersten Kapitel werden einige grundlegende Definitionen eingeführt sowie eine Reihe von allgemeinen Eigenschaften hergeleitet. Das zweite Kapitel stellt einen Zusammenhang zwischen Partitionen des Ereignisraumes und der abstandserzeugenden Funktion (AEF) $I(t)$ her. Csizsár [7] hat in einer Arbeit ein zum Hauptsatz dieses Kapitels äquivalentes Ergebnis erzielt; doch ist offenbar seine Beweisführung unrichtig. Das dritte Kapitel widmet sich der Frage, wie sich die Stützung von Wahrscheinlichkeitsmaßen auf die abstandserzeugende Funktion auswirkt. Die Suffizienz von Statistiken spielt u.a. in der Schätz- und Testtheorie eine herausragende Rolle. Die Bildmaße einer suffizienten Statistik, das ist das Hauptergebnis des vierten Kapitels, haben dieselbe abstandserzeugende Funktion wie die Ausgangsmaße. Csizsár [9] hat sich mit einer ähnlichen Fragestellung befaßt, jedoch mit der Einschränkung, daß die bedingten Erwartungen auch bedingte Wahrscheinlichkeitsverteilungen sind. (Siehe z.B. [39], S. 144, Satz 5.1.) Im nächsten Kapitel wird die Verträglichkeit der Produktbildung von Wahrscheinlichkeitsräumen mit der AEF untersucht. Für asymptotische Verfahren zieht man in der Statistik gern Grenzwertbetrachtungen heran. Im sechsten Kapitel werden deshalb Konvergenz von Wahrscheinlichkeitsmaßen und Konvergenz der abstandserzeugenden Funktion (in einem zu definierenden Sinne) gegenübergestellt. Abschließend werden im letzten Kapitel ein paar Anwendungen auf das Gebiet der parametrischen Testtheorie vorgeschlagen. Im Anhang befinden sich die abstandserzeugenden Funktionen für einige wichtige Wahrscheinlichkeitsmaße der Statistik.

I. Kapitel

Grundlegende Eigenschaften

Für das Folgende sei $(X, \mathcal{X})$ ein Maßraum, und P, Q seien Wahrscheinlichkeitsmaße auf $\mathcal{X}$, die durch ein σ -endliches Maß λ auf $\mathcal{X}$ dominiert seien (d.h. P und Q sind absolut stetig bez. λ). Mit

$$f := \frac{dP}{d\lambda} \quad (1.1)$$

$$g := \frac{dQ}{d\lambda} \quad (1.2)$$

bezeichnen wir die existierenden $[\lambda]$ - eindeutig bestimmten Radon-Nikodým-Dichten.

Für $E \in \mathcal{X}$ und $t \in [0, 1]$ definieren wir

$$I_E(P, Q)(t) := \int_E f(x)^t g(x)^{1-t} d\lambda(x) \quad (1.3)$$

Ist $E = X$, so setzen wir

$$I(P, Q)(t) := I_X(P, Q)(t) \quad (1.4)$$

$I(P, Q)(t)$ wird manchmal mit AEF bezeichnet (abstandserzeugende Funktion).

Für alle $E \in \mathcal{X}$ und alle $t \in [0, 1]$ existiert $I_E(P, Q)(t)$, und es ist

$$0 \leq I_E(P, Q)(t) \leq 1 \quad (1.5)$$

Wegen Kingman-Taylor [36] S. 183 gilt nämlich für $t \in [0, 1]$:

$$f^t g^{1-t} \leq tf + (1-t)g \quad (1.6)$$

Wenn man über E auf beiden Seiten integriert, so erhält man

$$0 \leq I_E(P, Q)(t) \leq tP(E) + (1-t)Q(E) \leq 1 \quad (1.7)$$

$I_E(P, Q)(t)$ ist unabhängig von der Wahl des dominierenden Maßes λ , denn wenn

$$\lambda_1 := P + Q \quad (1.8)$$

gilt:

$$\int_E f_1^t g_1^{1-t} d\lambda_1 = \int_E f^t g^{1-t} d\lambda \quad (1.9)$$

wobei

$$f_1 := \frac{dP}{d\lambda_1} \quad g_1 := \frac{dQ}{d\lambda_1} \quad (1.10)$$

Es ist offenbar $\lambda_1 \ll \lambda$, $P \ll \lambda_1$, $Q \ll \lambda_1$.

Wegen Halmos [20] S. 133, Theorem A, gilt dann

$$\frac{dP}{d\lambda} = \frac{dP}{d\lambda_1} \cdot \frac{d\lambda_1}{d\lambda} \quad [\lambda] \quad (1.11)$$

$$\frac{dQ}{d\lambda} = \frac{dQ}{d\lambda_1} \cdot \frac{d\lambda_1}{d\lambda} \quad [\lambda] \quad (1.12)$$

d.h. es ist

$$f = f_1 \cdot \frac{d\lambda_1}{d\lambda} \quad [\lambda] \quad (1.13)$$

$$g = g_1 \cdot \frac{d\lambda_1}{d\lambda} \quad [\lambda] \quad (1.14)$$

Mithin folgt

$$\begin{aligned} \int_E f^t g^{1-t} d\lambda &= \int_E \left[f_1 \cdot \frac{d\lambda_1}{d\lambda} \right]^t \left[g_1 \cdot \frac{d\lambda_1}{d\lambda} \right]^{1-t} d\lambda \\ &= \int_E f_1^t g_1^{1-t} \frac{d\lambda_1}{d\lambda} d\lambda \\ &= \int_E f_1^t g_1^{1-t} d\lambda_1 \end{aligned} \quad (1.15)$$

Zur letzten Identität in (1.15) vergleiche man Halmos [20] S. 134, Theorem B.

Es werden jetzt einige triviale Eigenschaften der Vollständigkeit halber aufgeführt. Wir setzen noch für eine Funktion

$$h: X \rightarrow \mathbb{R}$$

$$TR(h) := \{x | x \in X \wedge h(x) > 0\} \quad (1.16)$$

(Lies: Träger von h)

Es ist dann

$$I_E(P, Q)(t) = \int_{TR(f) \cap TR(g) \cap E} f^t g^{1-t} d\lambda \quad (1.17)$$

Also:

$$I_E(P, Q)(0) = Q(TR(f) \cap TR(g) \cap E) \quad (1.18)$$

$$I_E(P, Q)(1) = P(TR(f) \cap TR(g) \cap E) \quad (1.19)$$

$$I(P, Q)(0) = Q(TR(f) \cap TR(g)) \quad (1.20)$$

$$I(P, Q)(1) = P(TR(f) \cap TR(g)) \quad (1.21)$$

$$I(P, Q)\left(\frac{1}{2}\right) = \int \sqrt{fg} \, d\lambda \quad (1.22)$$

(Zu (1.22) vergleiche man Bhattacharya [5])

$$I_t(P||Q) = \frac{1}{t-1} \log I(P, Q)(t) \quad (1.23)$$

Dabei ist $I_t(P||Q)$ gerade das von Rényi [48] axiomatisch begründete Informationsmaß (bei Rényi heißt $I_t(P||Q)$ die relative Information der Ordnung t von P bez. Q , $t \neq 1$).

Ist $P \perp Q$, d.h. es gibt ein $M \in \mathcal{K}$ mit $P(M) = 0$, $Q(M) = 1$, so gilt für alle $E \in \mathcal{K}$ und alle $t \in [0, 1]$:

$$I_E(P, Q)(t) = 0 \quad (1.24)$$

Ist umgekehrt $I(P, Q)(t_0) = 0$ für ein $t_0 \in (0, 1)$, so gilt $P \perp Q$. Gilt nämlich $\int f^{t_0} g^{1-t_0} d\lambda = 0$, so ist $f^{t_0} g^{1-t_0} = 0$ λ -fast überall.

Sei $N := \{x | x \in X \wedge f^{t_0}(x) g^{1-t_0}(x) \neq 0\}$.

Dann ist $\lambda(N) = 0$ und $N = TR(f^{t_0}) \cap TR(g^{1-t_0})$.

Für M wähle man $TR(g)$. Weil $TR(f) = TR(f^{t_0})$ und $TR(g) = TR(g^{1-t_0})$ ist, folgt

$$\begin{aligned} P(TR(g)) &= P(TR(g) \setminus TR(f)) + P(TR(f) \cap TR(g)) \\ &= \int f \, d\lambda + P(TR(f) \cap TR(g)) \\ &\quad TR(g) \setminus TR(f) \\ &= 0 \end{aligned} \quad (1.25)$$

Hierbei wurde noch ausgenutzt, daß $P \ll \lambda$.

Trivialerweise ist $TR(g) \in \mathcal{K}$ und $Q(TR(g)) = 1$. Wird vorausgesetzt, daß $P \ll Q$ (d.h. Q dominiert P), so folgt $TR(f) \subset TR(g)$ [λ] und es ist für $E \in \mathcal{K}$:

$$I_E(P, Q)(0) = Q(TR(f) \cap E) \quad (1.26)$$

$$I_E(P, Q)(1) = P(E) \quad (1.27)$$

$$I(P, Q)(0) = Q(\text{TR}(f)) \quad (1.28)$$

$$I(P, Q)(1) = 1 \quad (1.29)$$

Ist sogar $P \equiv Q$ (d.h. $P \ll Q$ und $Q \ll P$), so folgt $\text{TR}(f) = \text{TR}(g) \quad [\lambda]$ und dann gilt:

$$I_E(P, Q)(0) = Q(E) \quad (1.30)$$

$$I_E(P, Q)(1) = P(E) \quad (1.31)$$

$$I(P, Q)(0) = 1 \quad (1.32)$$

$$I(P, Q)(1) = 1 \quad (1.33)$$

Es folgt der erste Satz.

SATZ 1.1

Für alle $E \in \mathcal{K}$ ist $I_E(P, Q)(\cdot)$ eine konvexe Funktion auf $[0, 1]$.

Beweis:

Seien $t_1, t_2 \in [0, 1]$, $\alpha_1, \alpha_2 \in [0, 1]$, $\alpha_1 + \alpha_2 = 1$.

Zu zeigen:

$$I_E(P, Q)(\alpha_1 t_1 + \alpha_2 t_2) \leq \alpha_1 I_E(P, Q)(t_1) + \alpha_2 I_E(P, Q)(t_2) \quad (1.34)$$

Nun ist nach Definition:

$$I_E(P, Q)(\alpha_1 t_1 + \alpha_2 t_2) = \int_E f^{\alpha_1 t_1 + \alpha_2 t_2} g^{1 - \alpha_1 t_1 - \alpha_2 t_2} d\lambda \quad (1.35)$$

$$= \int_E f^{\alpha_1 t_1} g^{-\alpha_1 t_1} f^{\alpha_2 t_2} g^{-\alpha_2 t_2} g^1 d\lambda \quad (1.36)$$

$$= \int_E \left[f^{t_1} g^{1-t_1} \right]^{\alpha_1} \left[f^{t_2} g^{1-t_2} \right]^{\alpha_2} d\lambda \quad (1.37)$$

wobei wir noch verwendeten, daß $1 = \alpha_1 + \alpha_2$.

Nun gilt aber (siehe Kingman-Taylor [36] S. 183):

$$[f^{t_1} g^{1-t_1}]^{\alpha_1} [f^{t_2} g^{1-t_2}]^{\alpha_2} \leq \alpha_1 f^{t_1} g^{1-t_1} + \alpha_2 f^{t_2} g^{1-t_2} \quad (1.38)$$

Danach ergibt sich

$$I_E(P, Q)(\alpha_1 t_1 + \alpha_2 t_2) \leq \alpha_1 \int_E f^{t_1} g^{1-t_1} d\lambda + \alpha_2 \int_E f^{t_2} g^{1-t_2} d\lambda \quad (1.39)$$

Das ist aber gerade die gewünschte Beziehung (1.34).

Bemerkung:

Sehr häufig liegt der Fall der zweimaligen Differenzierbarkeit von $I_E(P,Q)(\cdot)$ auf $[0,1]$ vor mit

$$\frac{d^2}{dt^2} I_E(P,Q)(t) = \int_E t^t g^{1-t} \left[\log \frac{f}{g} \right]^2 d\lambda \quad (1.40)$$

Dann ist die Konvexität von $I_E(P,Q)(\cdot)$ wegen $\frac{d^2}{dt^2} I_E(P,Q)(t) \geq 0$ noch einmal gezeigt. Man vergleiche etwa Kullback [40] S. 40. Prinzipiell kann jedoch, das zeigt der vorige Satz, auf die Differenzierbarkeit von $I_E(P,Q)(\cdot)$ verzichtet werden, wenn man die Konvexität nachweisen will.

Übrigens gibt es Wahrscheinlichkeitsmaße P, Q , so daß $I(P,Q)(\cdot)$ nicht überall differenzierbar ist. Dazu betrachten wir (siehe auch Kullback [40] S. 10):

$$X = \mathbb{N} \setminus \{1\}, \quad \mathcal{X} = \mathcal{P}(X)$$

$$P(x) = \frac{1}{k_1 x (\log x)^2} \quad \text{für } x \in X \quad (1.41)$$

$$Q(x) = \frac{1}{k_2 x^2 (\log x)^2} \quad \text{für } x \in X \quad (1.42)$$

wobei

$$k_1 = \sum_{n=2}^{\infty} \frac{1}{n(\log n)^2} \quad (1.43)$$

$$k_2 = \sum_{n=2}^{\infty} \frac{1}{n^2 (\log n)^2} \quad (1.44)$$

Wie man nachprüft, ist

$$I(P,Q)(t) = \left(\frac{1}{k_1} \right)^t \left(\frac{1}{k_2} \right)^{1-t} \sum_{x=2}^{\infty} \frac{1}{x^{2-t}} \cdot \frac{1}{(\log x)^2} \quad (1.45)$$

Leitet man $\sum_{x=2}^{\infty} \frac{1}{x^{2-t}} \cdot \frac{1}{(\log x)^2}$ formal nach t ab, so ergibt sich

$$\sum_{x=2}^{\infty} \frac{1}{x^{2-t}} \cdot \frac{1}{\log x} \cdot$$

Diese Reihe ist jedoch für $t = 1$ nicht konvergent (vgl. Knopp [37], S. 124), d.h. $I(P, Q)(\cdot)$ ist dort auch nicht differenzierbar. Dieser Sachverhalt ist besonders im Hinblick auf den Zusammenhang zwischen $I(P, Q)(\cdot)$ und dem Kullback-Leibler-Maß $K(P, Q)$ von Bedeutung. Mit der Konvexität von $I_E(P, Q)(\cdot)$ auf $[0, 1]$ ist auch die Stetigkeit auf $(0, 1)$ gezeigt (siehe Hewitt, Stromberg [23] S. 202).

Wir kommen jetzt zu weiteren Eigenschaften der AEF. Ist $P = Q$, so gilt für alle $E \in \mathcal{X}$ und alle $t \in [0, 1]$:

$$I_E(P, Q)(t) = P(E) = Q(E) \quad (1.46)$$

Interessanter ist die umgekehrte Richtung:

Ist $I(P, Q)(t_0) = 1$ für ein $t_0 \in (0, 1)$, so ist $P = Q$. Dazu benötigen wir das folgende

LEMMA 1.2

I: $[0, 1] \rightarrow [0, 1]$ sei eine konvexe Funktion. Es gebe $t_0 \in (0, 1)$ mit $I(t_0) = 1$. Dann gilt für alle $t \in [0, 1]$:

$$I(t) = 1 \quad (1.47)$$

Beweis:

Angenommen, es gäbe ein $t' \in [0, 1]$ mit $I(t') < 1$. OBdA sei $t_0 < t'$. Wir setzen $t'' := 0$. Da t_0 zwischen t'' und t' liegt, gibt es $\alpha_1, \alpha_2 > 0$ mit $\alpha_1 + \alpha_2 = 1$ und $t_0 = \alpha_1 t' + \alpha_2 t''$. Weil I eine konvexe Funktion ist, folgt:

$$I(\alpha_1 t' + \alpha_2 t'') \leq \alpha_1 I(t') + \alpha_2 I(t''), \text{ d.h.}$$

$$I(t_0) < \alpha_1 + \alpha_2 I(t''), \text{ und dann}$$

$$1 < \alpha_1 + \alpha_2 = 1$$

Das ist aber ein Widerspruch. Damit folgt die Behauptung.

Wegen Satz 1.1 und Lemma 1.2 folgt mithin, daß $I(P, Q)(t) = 1$ für alle $t \in [0, 1]$ gilt (sofern $I(P, Q)(t_0) = 1$ für ein $t_0 \in (0, 1)$). Dann ist auch $I(P, Q)(\frac{1}{2}) = 1$, und es folgt wegen Adhikari - Joshi [1] S. 63: $P = Q$.

$I(P, Q)(\cdot)$ kann als momenterzeugende Funktion der Zufallsvariablen $X = \log \frac{f}{g}$ bez. des Maßes Q aufgefaßt werden (vgl. etwa [55] S. 114), denn es ist

$$\begin{aligned}
 E_Q(e^{t \log \frac{f}{g}}) &= E_Q((\frac{f}{g})^t) \\
 &= \int f^t g^{-t} dQ \\
 &= \int f^t g^{1-t} d\lambda \\
 &= I(P, Q)(t)
 \end{aligned} \tag{1.48}$$

Hierbei bezeichnet E_Q den Erwartungsoperator bez. Q .

Aus den Eigenschaften der momenterzeugenden Funktion einer Zufallsvariablen folgt sofort

SATZ 1.3

$X_1, X_2 : X \rightarrow \mathbb{R}$ seien zwei unabhängige Zufallsvariablen (diskret oder stetig) mit den Dichtefunktionen $f(x_1, x_2)$ bzw. $g(x_1, x_2)$. $f_i(x_i)$, $g_i(x_i)$, $i = 1, 2$, seien die zugehörigen Randdichten. Dann gilt:

$$I(P, Q)(t) = I(P_1, Q_1)(t) \cdot I(P_2, Q_2)(t) \tag{1.49}$$

wobei P, Q, P_i, Q_i , $i = 1, 2$, die entsprechenden induzierten Bildmaße seien.

Wir kommen nun zu einigen in der Statistik und Wahrscheinlichkeitstheorie auftretenden Abstandsmaßen und stellen ein paar Relationen zu der betrachteten AEF her.

SATZ 1.4

Für alle $t \in [0, 1]$ gilt:

$$I(P, Q)(t) \geq 1 - \sup_{E \in \mathcal{X}} |P(E) - Q(E)| \tag{1.50}$$

Beweis:

Er folgt direkt aus nachstehender Kette von Gleichungen und Ungleichungen:

$$\begin{aligned}
 I(P, Q)(t) &= \int f^t g^{1-t} d\lambda \\
 &= \int_{\{f > g\}} f^t g^{1-t} d\lambda + \int_{\{f \leq g\}} f^t g^{1-t} d\lambda
 \end{aligned}$$

$$\begin{aligned}
 &= \int_{\{f > g\}} \left[\frac{f(x)}{g(x)} \right]^t dQ(x) + \int_{\{f \leq g\}} \left[\frac{g(x)}{f(x)} \right]^{1-t} dP(x) \\
 &\geq Q(f > g) + 1 - P(f > g) \\
 &= 1 - \frac{1}{2} \int |f - g| d\lambda \\
 &= 1 - \sup_{E \in \mathcal{X}} |P(E) - Q(E)| \quad (1.51)
 \end{aligned}$$

Zu den beiden letzten Ungleichungen vergleiche man auch Adhikari - Joshi [1] S. 71 u. 73.

Sind $\{P_n\}_{n \in \mathbb{N}}$, $\{Q_n\}_{n \in \mathbb{N}}$ zwei Folgen von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$, und setzt man

$$d(P_n, Q_n) := \sup_{E \in \mathcal{X}} |P_n(E) - Q_n(E)| \quad (1.52)$$

so gilt folgende Implikation für alle $t \in [0, 1]$:

$$\lim_{n \rightarrow \infty} d(P_n, Q_n) = 0 \Rightarrow \lim_{n \rightarrow \infty} I(P_n, Q_n)(t) = 1 \quad (1.53)$$

Dies folgt direkt aus Satz 1.4.

Offensichtlich ist für $t \in (0, 1)$:

$$\lim_{n \rightarrow \infty} I_t(P_n || Q) = 0 \Leftrightarrow \lim_{n \rightarrow \infty} I(P_n, Q)(t) = 1 \quad (1.54)$$

(Man vergleiche die Definition auf S. 8.)

Mit der Arbeit von Csizsár [7] ergibt sich der

SATZ 1.5

$\{P_n\}$ sei eine Folge von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$. Dann sind die folgenden Aussagen äquivalent für $t \in (0, 1)$:

$$(i) \quad \lim_{n \rightarrow \infty} I(P_n, Q)(t) = 1 \quad (1.55)$$

$$(ii) \quad \lim_{n \rightarrow \infty} I_t(P_n || Q) = 0 \quad (1.56)$$

$$(iii) \quad \lim_{n \rightarrow \infty} d(P_n, Q) = 0 \quad (1.57)$$

Insbesondere ist mit $f_n := \frac{dP}{d\lambda}$ die Konvergenz von f_n gegen g im Mittel (das ist die Aussage von (1.57)) mit der Aussage (1.55) äquivalent.

Auch zu dem Kullback-Leibler-Maß [40]

$$K(P, Q) = \int f \log \frac{f}{g} d\lambda \quad (1.58)$$

hat die AEF interessante Verbindungen. Dazu brauchen wir das folgende Lemma:

LEMMA 1.6

$(X, \mathcal{X}, W)$ sei ein Wahrscheinlichkeitsraum

$$h: X \rightarrow \mathbb{R}^+ \cup \{0\}$$

sei W -integrierbar. Setzt man $\log 0 = -\infty$, so gilt:

$$\int \log h dW \leq \log \int h dW \quad (1.59)$$

Beweis:

Siehe Hewitt-Stromberg [23] S.201.

Wir können jetzt formulieren

SATZ 1.7

Für alle $t \in [0, 1]$ sei $I(P, Q)(t) \neq 0$.

(i) Für alle $t \in [0, 1)$ ist dann

$$(1 - t)K(P, Q) \geq \log \frac{1}{I(P, Q)(t)} \quad (1.60)$$

(ii) Ist zusätzlich $K(P, Q) < \infty$, so gilt (1.60) sogar für alle $t \in [0, 1]$.

Beweis:

(i)

I. Fall $K(P, Q) = \infty$ fertig

II. Fall $K(P, Q) < \infty$

Wir setzen im vorstehenden Lemma 1.6

$$h(x) = \begin{cases} \left[\frac{f(x)}{I(P, Q)} \right]^{1-t} & \text{falls } f(x) \neq 0 \\ 0 & \text{falls } f(x) = 0 \end{cases} \quad (1.61)$$

Es ist dann

$$\begin{aligned} \int h \, dP &= \int \left(\frac{g}{f} \right)^{1-t} f \, d\lambda \\ &= I(P, Q)(t) \\ &\leq 1 \end{aligned} \quad (1.62)$$

und damit ist h P -integrabel.

Somit folgt wegen Lemma 1.6:

$$\int \log \left(\frac{g}{f} \right)^{1-t} f \, d\lambda \leq \log I(P, Q)(t) \quad (1.63)$$

Nutzt man die Eigenschaften des Logarithmus aus, erhält man

$$-(1-t)K(P, Q) \leq \log I(P, Q)(t) \quad (1.64)$$

oder

$$(1-t)K(P, Q) \geq \log \frac{1}{I(P, Q)(t)} \quad (1.65)$$

(ii)

Analog zu (i), II. Fall.

Und nun zu einigen Eigenschaften, die sich aus der Differenzierbarkeit von $I(P, Q)(\cdot)$ ergeben.

Ist $I_E(P, Q)(\cdot)$ in einer Umgebung von $t_0 \in [0, 1]$ n -mal differenzierbar, und sind die Operatoren $\int$ und $\frac{d}{dt}$ vertauschbar, so gilt:

$$\frac{d^n}{dt^n} I_E(P, Q)(t) \Big|_{t=t_0} = \int_E f(x)^{t_0} g(x)^{1-t_0} \left[\log \frac{f(x)}{g(x)} \right]^n d\lambda(x) \quad (1.66)$$

Über die Vertauschbarkeit von $\int$ und $\frac{d}{dt}$ lese man etwa McShane [57] S. 216 und ferner Fichtenholz [14] Bd. II S. 454 und S. 736.

Dann ist auch

$$\frac{d}{dt} I(P, Q)(t) \Big|_{t=1} = K(P, Q) \quad (1.67)$$

$$\frac{d}{dt} I(P, Q)(t) \Big|_{t=0} = -K(Q, P) \quad (1.68)$$

und vermöge de l'Hospital:

$$\lim_{t \rightarrow 1} \frac{1}{t-1} \log I(P, Q)(t) = K(P, Q) \quad (1.69)$$

d.h.

$$\lim_{t \rightarrow 1} I_t(P||Q) = K(P,Q) \quad (1.70)$$

Mit Hilfe von (1.67) und (1.68) erhält man dann auch die Kullback-Leibler-Divergenz [40] S. 6:

$$D(P,Q) = K(P,Q) + K(Q,P) \quad (1.71)$$

Bevor wir nun zu dem wichtigsten Satz des Kapitels kommen, sei noch eine andere wichtige Beziehung hervorgehoben. Ist $E \in \mathcal{X}$ und $t \in [0,1]$, so gilt:

$$I_E(P,Q)(t) \leq P(E)^t Q(E)^{1-t} \quad (1.72)$$

Dies folgt sofort aus der bekannten Hölderschen Ungleichung (vgl. Bauer [2] S. 60 mit $\frac{1}{p} = t, \frac{1}{q} = 1-t$). Interessant ist diese Ungleichung im Zusammenhang mit Partitionen des Raumes X . Eine Familie

$\mathcal{I} = \{E_i\}_{i \in I}$ heißt dabei Partition von X , wenn gilt:

$$E_i \cap E_j = \emptyset \quad \text{für } i \neq j \quad (1.73)$$

$$\bigcup_{i \in I} E_i = X \quad (1.74)$$

$$E_i \in \mathcal{X} \quad \text{für } i \in I \quad (1.75)$$

$\mathcal{I}$ heißt abzählbar unendliche Partition, falls I abzählbar unendlich ist.

$\mathcal{I}$ heißt endliche Partition, falls I endlich ist.

Ist $\mathcal{I} = \{E_i\}_{i \in I}$ eine abzählbar unendliche oder endliche Partition, so hat man

$$I(P,Q)(t) \leq \sum_{i \in I} P(E_i)^t Q(E_i)^{1-t} \quad (1.76)$$

Dies folgt direkt aus nachstehender Identität und (1.77)

$$I(P,Q)(t) = \sum_{i \in I} I_{E_i}(P,Q)(t)$$

Für den bereits angekündigten wichtigen Satz braucht man noch einige Definitionen.

Ist

$$T : X \rightarrow Y$$

eine $(\mathcal{X}, \mathcal{A})$ -meßbare Transformation, wobei $(Y, \mathcal{B})$ ein Meßraum ist,

so kann man das durch T induzierte Bildmaß $T(P)$ auf $\mathcal{W}$ betrachten, das durch

$$T(P)(F) := P(T^{-1}(F)) \quad (1.78)$$

mit $F \in \mathcal{W}$ erklärt ist. Analog bildet man $T(Q)$ und $T(\lambda)$. Wie man z.B. Halmos [20] S. 229 entnehmen kann, gilt:

$$T(P) \ll T(\lambda), \quad T(Q) \ll T(\lambda) \quad (1.79)$$

Setzt man

$$E_\lambda(f|T=y) := \frac{dT(P)}{dT(\lambda)}(y) \quad (1.80)$$

$$E_\lambda(g|T=y) := \frac{dT(Q)}{dT(\lambda)}(y) \quad (1.81)$$

so ist für $F \in \mathcal{W}$:

$$\int_{T^{-1}(F)} f(x) d\lambda(x) = \int_F E_\lambda(f|T=y) dT(\lambda)(y) \quad (1.82)$$

$$\int_{T^{-1}(F)} g(x) d\lambda(x) = \int_F E_\lambda(g|T=y) dT(\lambda)(y) \quad (1.83)$$

Und jetzt der

SATZ 1.8

$(Y, \mathcal{W})$ sei ein Maßraum, und

$$T : X \rightarrow Y$$

sei $(\mathcal{X}, \mathcal{W})$ -meßbar. Dann gilt für alle $t \in [0, 1]$:

$$I(P, Q)(t) \leq I(T(P), T(Q))(t) \quad (1.84)$$

Beweis:

Es genügt, (1.84) für alle $t \in (0, 1)$ nachzuweisen. Hierzu zeigen wir also für $t \in (0, 1)$:

$$\int f(x)^t g(x)^{1-t} d\lambda(x) \leq \int E_\lambda(f|T=y)^t E_\lambda(g|T=y)^{1-t} dT(\lambda)(y) \quad (1.85)$$

Man betrachte die Funktion

$$h : X \rightarrow \mathbb{R}$$

$$\text{mit } x \rightarrow h(x) = \begin{cases} \frac{E_{\lambda}(g|T(x))}{E_{\lambda}(f|T(x))} & \text{falls } E_{\lambda}(f|T(x)) \neq 0 \\ 0 & \text{sonst} \end{cases} \quad (1.86)$$

Bildet man die σ -Algebra

$$T^{-1}(\mathcal{W}) := \{T^{-1}(F) | F \in \mathcal{W}\} \quad (1.87)$$

und bezeichnet man mit $\mathcal{L}^n$ die σ -Algebra der Borelschen Mengen des $\mathbb{R}^n$, so ist h $(T^{-1}(\mathcal{W}) - \mathcal{L}^1)$ -meßbar. Dies folgt aus zwei Fakten, nämlich:

- (i) T ist $(T^{-1}(\mathcal{W}) - \mathcal{W})$ -meßbar
- (ii) $E_{\lambda}(f|\cdot)$ und $E_{\lambda}(g|\cdot)$ sind $(\mathcal{W} - \mathcal{L}^1)$ -meßbar.

Für jedes $n \in \mathbb{N}$ und jedes $m \in \mathbb{N} \cup \{0\}$ sei definiert

$$I_{mn} := \left(\frac{m}{2^n}, \frac{m+1}{2^n} \right] \quad (1.88)$$

Da die Funktion

$$\psi : \mathbb{R}^+ \cup \{0\} \rightarrow \mathbb{R}$$

$$\text{mit } z \rightarrow \psi(z) = z^{1-t}$$

stetig ist, ist h^{1-t} $(T^{-1}(\mathcal{W}) - \mathcal{L}^1)$ -meßbar.

Setzt man

$$E_{mn} := \{x | x \in X \wedge h^{1-t}(x) \in I_{mn}\} \quad (1.89)$$

so ist $E_{mn} \in T^{-1}(\mathcal{W})$.

Für jedes $n \in \mathbb{N}$ gilt außerdem

$$X = \bigcup_{m=0}^{\infty} E_{mn} \quad (1.90)$$

Für $n \in \mathbb{N}$ definieren wir nun weiter:

$$u_n = \sum_{m=0}^{\infty} \frac{m}{2^n} P(E_{mn}) \quad (1.91)$$

Wegen des Resultates bei Halmos [20] S. 115 hat man

$$\lim_{n \rightarrow \infty} u_n = \int h(x)^{1-t} dP(x)$$

$$\begin{aligned}
 &= \int \left[\frac{E_{\lambda}(g|T(x))}{E_{\lambda}(f|T(x))} \right]^{1-t} dP(x) \\
 &= \int E_{\lambda}(f|T(x))^{t-1} E_{\lambda}(g|T(x))^{1-t} dP(x) \\
 &= \int E_{\lambda}(f|T=y)^{t-1} E_{\lambda}(g|T=y)^{1-t} dT(P)(y) \quad (1.92)
 \end{aligned}$$

Nun ist

$$\frac{dT(P)}{dT(\lambda)} = E_{\lambda}(f|\cdot) \quad (1.93)$$

und somit

$$\begin{aligned}
 \lim_{n \rightarrow \infty} u_n &= \int E_{\lambda}(f|T=y)^t E_{\lambda}(g|T=y)^{1-t} dT(\lambda)(y) \\
 &= I(T(P), T(Q))(t) \quad (1.94)
 \end{aligned}$$

Ist jedoch $x \in E_{mn}$, so erhält man

$$\frac{m}{2^n} < h(x)^{1-t} \leq \frac{m+1}{2^n} \quad (1.95)$$

d.h.

$$\left[\frac{m}{2^n} \right]^{1-t} < h(x) \leq \left[\frac{m+1}{2^n} \right]^{1-t} \quad (1.96)$$

Multipliziert man auf allen Seiten mit $\chi_{E_{mn}}$ (der Indikatorfunktion von E_{mn}), so ergibt sich nach Integration:

$$\left[\frac{m}{2^n} \right]^{1-t} P(E_{mn}) \leq \int_{E_{mn}} h(x) dP(x) \quad (1.97)$$

und

$$\int_{E_{mn}} h(x) dP(x) \leq \left[\frac{m+1}{2^n} \right]^{1-t} P(E_{mn}) \quad (1.98)$$

Nun ist

$$\int_{E_{mn}} h(x) dP(x) = \int_{E_{mn}} \frac{E_{\lambda}(g|T(x))}{E_{\lambda}(f|T(x))} dP(x) \quad (1.99)$$

Weil $E_{mn} \in T^{-1}(\mathcal{W})$ ist, gibt es ein $F_{mn} \in \mathcal{W}$ mit

$$E_{mn} = T^{-1} [F_{mn}] \quad (1.100)$$

Also hat man

$$\begin{aligned} \int_{E_{mn}} h(x) dP(x) &= T^{-1} \left[F_{mn} \right] \frac{E_{\lambda}(g|T(x))}{E_{\lambda}(f|T(x))} dP(x) \\ &= \int_{F_{mn}} \frac{E_{\lambda}(g|T=y)}{E_{\lambda}(f|T=y)} dT(P)(y) \end{aligned} \quad (1.101)$$

Hierzu vergleiche man Halmos [20] S. 228.

Da

$$E_{\lambda}(f|\cdot) = \frac{dT(P)}{dT(\lambda)} \quad \text{gilt, so folgt}$$

$$\begin{aligned} \int_{E_{mn}} h(x) dP(x) &= \int_{F_{mn}} E_{\lambda}(g|T=y) dT(\lambda)(y) \\ &= T(Q)(F_{mn}) \\ &= Q(E_{mn}) \end{aligned} \quad (1.102)$$

Setzt man oben ein, so ist

$$\left[\frac{m}{2^n} \right]^{1-t} P(E_{mn}) \leq Q(E_{mn}) \quad (1.103)$$

und

$$Q(E_{mn}) \leq \left[\frac{m+1}{2^n} \right]^{1-t} P(E_{mn}) \quad (1.104)$$

Insgesamt ergibt sich nach einigen elementaren Umformungen:

$$\frac{m}{2^n} P(E_{mn}) \leq P(E_{mn})^t Q(E_{mn})^{1-t} \leq \left(\frac{m}{2^n} + \frac{1}{2^n} \right) P(E_{mn}) \quad (1.105)$$

Summation über m liefert jetzt für jedes $n \in \mathbb{N}$:

$$u_n \leq \sum_{m=0}^{\infty} P(E_{mn})^t Q(E_{mn})^{1-t} \leq u_n + \frac{1}{2^n} \quad (1.106)$$

Damit gilt

$$\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \sum_{m=0}^{\infty} P(E_{mn})^t Q(E_{mn})^{1-t} \quad (1.107)$$

Andererseits ist für $n \in \mathbb{N}$:

$$I(P, Q)(t) \leq \sum_{m=0}^{\infty} P(E_{mn})^t Q(E_{mn})^{1-t} \quad (1.108)$$

Mithin muß gelten:

$$I(P, Q)(t) \leq \lim_{n \rightarrow \infty} u_n \quad (1.109)$$

Oben wurde jedoch bewiesen, daß

$$\lim_{n \rightarrow \infty} u_n = I(T(P), T(Q))(t), \quad (1.110)$$

Somit ist alles gezeigt.

Volle Bedeutung erlangt der vorstehende Satz erst in den folgenden Kapiteln. Es soll hier nur noch ein naheliegendes Korollar angegeben werden.

KOROLLAR 1.9

Sei $\mathcal{X}'$ eine σ -Algebra auf X und sei $\mathcal{X}' \subset \mathcal{X}$. P' bzw. Q' seien die Einschränkungen von P und Q auf $\mathcal{X}'$. Dann gilt für alle $t \in [0, 1]$:

$$I(P, Q)(t) \leq I(P', Q')(t) \quad (1.111)$$

Beweis:

Wir betrachten die Transformation

$$1_X : X \rightarrow X$$

(1_X Identität auf X)

Da $\mathcal{X}' \subset \mathcal{X}$, ist 1_X $(\mathcal{X}, \mathcal{X}')$ -meßbar. Wegen Satz 1.8 gilt für alle $t \in [0, 1]$:

$$I(P, Q)(t) \leq I(1_X(P), 1_X(Q))(t)$$

Nun ist $1_X(P) = P'$ und $1_X(Q) = Q'$, und die Behauptung folgt sofort.

II. Kapitel

Ein Hauptsatz

Im vorigen Kapitel wurde gezeigt, daß für jede endliche Partition

$$\mathcal{T} = \{E_i\}_{i=1}^n \text{ gilt:}$$

$$I(P, Q)(t) \leq \sum_{i=1}^n P(E_i)^t Q(E_i)^{1-t} \quad (2.1)$$

Man kann hier vermuten, daß bei "Verfeinerung" der Partition die rechtsstehende Summe näher an $I(P, Q)(t)$ heranwächst. Dies trifft tatsächlich zu, wie das nächste Lemma zeigt. Zunächst jedoch die folgende

Definition:

$\mathcal{T}$, $\mathcal{T}'$ seien Partitionen von X . $\mathcal{T}'$ heißt feiner als $\mathcal{T}$ (in Zeichen: $\mathcal{T} < \mathcal{T}'$), falls gilt: Zu jedem $E \in \mathcal{T}$ gibt es ein $\mathcal{T}'_E \subset \mathcal{T}'$ mit

$$E = \bigcup_{E' \in \mathcal{T}'_E} E'$$

Anschaulich heißt dies, daß sich jedes $E \in \mathcal{T}$ mit Elementen $E' \in \mathcal{T}'$ "auspflastern" läßt.

LEMMA 2.1

$\mathcal{T} = \{E_i\}_{i=1}^n$, $\mathcal{T}' = \{E'_j\}_{j=1}^k$ seien endliche Partitionen von X , $\mathcal{T} < \mathcal{T}'$. Dann gilt für alle $t \in [0, 1]$:

$$\sum_{j=1}^k P(E'_j)^t Q(E'_j)^{1-t} \leq \sum_{i=1}^n P(E_i)^t Q(E_i)^{1-t} \quad (2.2)$$

Beweis:

Sei $E_i \in \mathcal{T}$ beliebig. Da $\mathcal{T}'$ feiner als $\mathcal{T}$ ist, gibt es paarweise disjunkte Mengen $E'_{ji} \in \mathcal{T}'$, $j=1, \dots, m_i$ und

$$E_i = \bigcup_{j=1}^{m_i} E'_{ji} \quad (2.3)$$

Diese Darstellung ist für alle $E_i \in \mathcal{T}$ möglich. Mithin ist

$$\sum_{i=1}^n P(E_i)^t Q(E_i)^{1-t} = \sum_{i=1}^n \left(\sum_{j=1}^{m_i} P(E'_{ji}) \right)^t \cdot \left(\sum_{j=1}^{m_i} Q(E'_{ji}) \right)^{1-t} \quad (2.4)$$

Nun gilt aber wegen Hardy-Littlewood-Pólya [22] S. 22:

$$\sum_{j=1}^{m_i} P(E'_{ji})^t Q(E'_{ji})^{1-t} \leq \left(\sum_{j=1}^{m_i} P(E'_{ji}) \right)^t \left(\sum_{j=1}^{m_i} Q(E'_{ji}) \right)^{1-t} \quad (2.5)$$

So ergibt sich

$$\sum_{i=1}^n \sum_{j=1}^{m_i} P(E'_{ji})^t Q(E'_{ji})^{1-t} \leq \sum_{i=1}^n P(E_i)^t Q(E_i)^{1-t} \quad (2.6)$$

Weil $\mathcal{I}$ und $\mathcal{I}'$ endliche Partitionen sind, gilt notwendig:

$$\sum_{i=1}^n m_i = k \quad (2.7)$$

und dann

$$\sum_{j=1}^k P(E'_j)^t Q(E'_j)^{1-t} = \sum_{i=1}^n \sum_{j=1}^{m_i} P(E'_{ji})^t Q(E'_{ji})^{1-t} \quad (2.8)$$

Das ist aber auch alles, was zu zeigen war.

Wir betrachten jetzt noch einige

Beispiele:

(i) Wähle $X = \mathbb{N} \cup \{0\}$, $\mathcal{X} = \mathcal{P}(\mathbb{N} \cup \{0\})$ und auf $\mathcal{X}$ die beiden Pascalverteilungen:

$$P(x) = (1-p)p^x \quad 0 < p < 1 \quad (2.9)$$

$$Q(x) = (1-q)q^x \quad 0 < q < 1 \quad (2.10)$$

mit $x \in X$.

Für $\mathcal{I}$ wähle man $\{E_i\}_{i=1}^2$ mit

$$E_1 = \{0\}, \quad E_2 = \mathbb{N} \quad (2.11)$$

Für $\mathcal{I}'$ wähle man $\{E'_j\}_{j=1}^3$ mit

$$E'_1 = \{0\}, \quad E'_2 = \{1\}, \quad E'_3 = \mathbb{N} \setminus \{1\} \quad (2.12)$$

Wie man sofort erkennt, ist

$$\begin{aligned} P(E_1) &= P(E'_1) = 1-p & ; & & Q(E_1) &= Q(E'_1) = 1-q \\ P(E_2) &= p & ; & & Q(E_2) &= q \end{aligned}$$

$$\begin{aligned} P(E'_2) &= (1-p)p & ; & & Q(E'_2) &= (1-q)q \\ P(E'_3) &= p^2 & ; & & Q(E'_3) &= q^2 \end{aligned} \quad (2.13)$$

und danach

$$\sum_{i=1}^2 P(E'_i)^t Q(E'_i)^{1-t} = (1-p)^t (1-q)^{1-t} + p^t q^{1-t} \quad (2.14)$$

$$\begin{aligned} \sum_{j=1}^3 P(E'_j)^t Q(E'_j)^{1-t} &= (1-p)^t (1-q)^{1-t} \\ &+ [(1-p)p]^t [(1-q)q]^{1-t} + p^{2t} q^{2(1-t)} \end{aligned} \quad (2.15)$$

Also insgesamt für alle $t \in [0,1]$:

$$\begin{aligned} (1-p)^t (1-q)^{1-t} + [(1-p)p]^t [(1-q)q]^{1-t} + p^{2t} q^{2(1-t)} &\leq \\ (1-p)^t (1-q)^{1-t} + p^t q^{1-t} \end{aligned} \quad (2.16)$$

(ii) Es kann natürlich auch der Fall der Gleichheit auftreten, wie man im folgenden sieht:

$X = [R, \mathcal{X}] = \mathcal{X}^1$. Auf $\mathcal{X}$ betrachten wir die beiden Wahrscheinlichkeitsmaße P, Q mit jeweiliger Rechteckverteilung

$$\begin{aligned} f(x) &= \begin{cases} 1 & \text{falls } x \in [0,1] \\ 0 & \text{sonst} \end{cases} \\ g(x) &= \begin{cases} \frac{1}{2} & \text{falls } x \in [0,2] \\ 0 & \text{sonst} \end{cases} \end{aligned} \quad (2.17)$$

Für $\mathcal{I}$ sei gewählt $\{E_1, E_2\}$ mit

$$E_1 = (-\infty, 1) \quad ; \quad E_2 = [1, \infty) \quad (2.18)$$

Für $\mathcal{I}'$ wählen wir $\{E'_1, E'_2, E'_3, E'_4\}$ mit

$$E'_1 = (-\infty, \frac{1}{2}) \quad ; \quad E'_2 = [\frac{1}{2}, 1) \quad ; \quad E'_3 = [1, 1\frac{1}{2}) \quad ; \quad E'_4 = [1\frac{1}{2}, \infty) \quad (2.19)$$

Man erhält schnell:

$$\begin{aligned} P(E_1) &= 1, & P(E_2) &= 0, & Q(E_1) &= \frac{1}{2}, & Q(E_2) &= \frac{1}{2} \\ P(E'_1) &= \frac{1}{2}, & P(E'_2) &= \frac{1}{2}, & P(E'_3) &= 0, & P(E'_4) &= 0 \\ Q(E'_1) &= \frac{1}{4}, & Q(E'_2) &= \frac{1}{4}, & Q(E'_3) &= \frac{1}{4}, & Q(E'_4) &= \frac{1}{4} \end{aligned} \quad (2.20)$$

sowie

$$\sum_{i=1}^2 P(E_i)^t Q(E_i)^{1-t} = \left(\frac{1}{2}\right)^{1-t} \quad (2.21)$$

$$\sum_{j=1}^4 P(E_j)^t Q(E_j)^{1-t} = \left(\frac{1}{2}\right)^{1-t} \quad (2.22)$$

Bemerkung:

Man beachte, daß in diesem Fall

$$I(P,Q)(t) = \left(\frac{1}{2}\right)^{1-t} \quad (2.23)$$

ist !

Dem Lemma 2.1 zufolge wird jede Verfeinerung der Partition von X eine bessere Approximation von $I(P,Q)(t)$ bewirken (durch Summen der Form $\sum P(E_i)^t Q(E_i)^{1-t}$). Dies begründet den Verdacht, daß $I(P,Q)(t)$ sich als Infimum gewisser Summen darstellen läßt. Beim Vortasten in dieser Richtung wird man jedoch bald erkennen, daß man ohne Einschränkungen im Hinblick auf die Struktur von $\mathcal{X}$ nicht auskommen wird. Diese Restriktionen fallen so mild aus (Separabilität von $\mathcal{X}$), daß man immer noch mit dem Grad der Allgemeinheit zufrieden sein kann.

Dem eigentlichen zentralen Satz dieses Kapitels stellen wir nun eine Reihe von Definitionen und Hilfssätzen voran.

Zunächst die Definitionen:

A, B seien Mengen. Dann ist

$$A \Delta B := (A \setminus B) \cup (B \setminus A) \quad (2.24)$$

die symmetrische Differenz von A und B .

$\mathcal{E}$ sei ein System von Teilmengen von X . Dann ist $\sigma(\mathcal{E})$ die von $\mathcal{E}$ erzeugte σ -Algebra auf X (die kleinste σ -Algebra auf X , die $\mathcal{E}$ umfaßt).

$\mathcal{O}$ sei ein System von Teilmengen von X .

$\mathcal{O}$ heißt Algebra auf X , wenn gilt:

- (i) $\emptyset \in \mathcal{O}$, $X \in \mathcal{O}$
- (ii) $A, B \in \mathcal{O}$, dann auch $A \setminus B \in \mathcal{O}$
- (iii) $A, B \in \mathcal{O}$, dann auch $A \cup B \in \mathcal{O}$

$\{\mathcal{O}_n\}_{n \in \mathbb{N}}$ sei eine aufsteigende Folge von σ -Algebren auf X , λ ein Wahrscheinlichkeitsmaß auf $\mathcal{X}$, $\mathcal{O}_n \subset \mathcal{X}$ für alle $n \in \mathbb{N}$.

$\phi_n : X \rightarrow \mathbb{R}$ sei eine Folge P -integrabler, $(\mathcal{U}_n - \mathcal{L}^1)$ -meßbarer Funktionen. $\{\mathcal{U}_n, \phi_n\}_{n \in \mathbb{N}}$ heißt Martingal, wenn für alle $n \in \mathbb{N}$ gilt

$$E_P(\phi_{n+1} | \mathcal{U}_n) = \phi_n \quad [P] \quad (2.25)$$

Es folgen die Hilfssätze.

LEMMA 2.2

(Vgl. Doob [11], S. 606 Supplement § 2, Theorem 2.4)

Sei $\mathcal{F}_0$ eine Algebra auf X , $\mathcal{F}_1$ eine σ -Algebra auf X mit $\sigma(\mathcal{F}_0) \subset \mathcal{F}_1$. m sei ein Maß auf $\mathcal{F}_1$ mit der Eigenschaft

$$\begin{array}{ccc} \bigwedge & \bigvee & \\ A & B & \\ A \in \mathcal{F}_1 & B \in \mathcal{F}_0 & \end{array} \quad m(A \Delta B) = 0 \quad (2.26)$$

Dann gilt für jedes $(\mathcal{F}_1 - \mathcal{L}^1)$ -meßbare, m -integrierte

$$h : X \rightarrow \mathbb{R},$$

daß

$$\begin{array}{ccc} \bigwedge & \bigvee & \\ \epsilon & x_\epsilon & \int |h - x_\epsilon| \, dm < \epsilon \end{array} \quad (2.27)$$

$$\epsilon > 0 \quad x_\epsilon : X \rightarrow \mathbb{R}$$

$$x_\epsilon = \sum_{i=1}^n \alpha_i x_{B_i}$$

$$B_i \in \mathcal{F}_0$$

LEMMA 2.3

(Vgl. Witting [56], Kap. 2.3, Def. 2.12 und Hilfssatz 2.13 S. 61.)

$\mathcal{F}_1$ sei separable σ -Algebra auf X (d.h. $\mathcal{F}_1 = \sigma(\mathcal{E})$, $\mathcal{E}$ abzählbar). Dann gibt es eine abzählbare Algebra $\mathcal{F}_0 \subset \mathcal{F}_1$, so daß für jedes endliche Maß m auf $\mathcal{F}_1$ die Voraussetzung des Lemmas 2.2 erfüllt ist.

LEMMA 2.4

(Vgl. Bauer [2], S. 83, Satz 19.3)

sei ein endliches Maß auf $\mathcal{X}$. Konvergiert eine Folge $\{h_n\}_{n \in \mathbb{N}}$

von Funktionen aus $L^1(\mu)$ im Mittel gegen eine Funktion h aus $L^1(\mu)$, d.h. es gilt:

$$\lim_{n \rightarrow \infty} \int |h_n - h| d\mu = 0 \quad (2.28)$$

so konvergiert $\{h_n\}_{n \in \mathbb{N}}$ auch μ -stochastisch gegen h , d.h. für alle $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} \mu(\{x \in X \wedge |h_n(x) - h(x)| \geq \varepsilon\}) = 0 \quad (2.29)$$

LEMMA 2.5

(Vgl. Bauer [2], S. 85, Satz 19.6)

μ sei ein endliches Maß auf $\mathcal{X}$. Für jedes $n \in \mathbb{N}$ sei

$$h_n : X \rightarrow \mathbb{R}$$

eine $(\mathcal{X} - \mathcal{L}^1)$ -meßbare Funktion, $\{h_n\}_{n \in \mathbb{N}}$ konvergiere μ -stochastisch gegen eine $(\mathcal{X} - \mathcal{L}^1)$ -meßbare Funktion

$$h : X \rightarrow \mathbb{R}$$

Dann kann aus $\{h_n\}_{n \in \mathbb{N}}$ eine Teilfolge $\{h_{n_k}\}_{k \in \mathbb{N}}$ ausgesondert werden,

die μ -fast überall gegen h konvergiert, d.h. für $x \in X$ gilt

$$\lim_{k \rightarrow \infty} h_{n_k}(x) = h(x) \quad [\mu] \quad (2.30)$$

LEMMA 2.6

(Vgl. Bauer [2], S. 83, Satz 19.4)

μ sei ein endliches Maß auf $\mathcal{X}$. Für jedes $n \in \mathbb{N}$ sei

$$h_n : X \rightarrow \mathbb{R}$$

eine $(\mathcal{X} - \mathcal{L}^1)$ -meßbare Funktion, die μ -fast überall gegen eine $(\mathcal{X} - \mathcal{L}^1)$ -meßbare Funktion

$$h : X \rightarrow \mathbb{R}$$

konvergiere. Dann konvergiert $\{h_n\}_{n \in \mathbb{N}}$ auch μ -stochastisch gegen h .

LEMMA 2.7

(Vgl. Lemmata 2.4 und 2.5.)

μ sei ein endliches Maß auf $\mathcal{X}$. Konvergiert eine Folge $\{h_n\}_{n \in \mathbb{N}}$ von Funktionen aus $L^1(\mu)$ im Mittel gegen eine Funktion aus $L^1(\mu)$, so kann aus $\{h_n\}_{n \in \mathbb{N}}$ eine Teilfolge $\{h_{n_k}\}_{k \in \mathbb{N}}$ ausgesondert werden, die μ -fast überall gegen h konvergiert.

LEMMA 2.8

(Vgl. Lemmata 2.3, 2.7 und Witting [56], Hilfssätze 2.13, 2.16.)

$(X, \mathcal{F}_1)$ sei Maßraum mit separablem $\mathcal{F}_1 \cdot \mathcal{F}_0$ sei eine abzählbare Algebra mit $\sigma(\mathcal{F}_0) = \mathcal{F}_1$

$$h : X \rightarrow \mathbb{R}^+ \cup \{0\}$$

sei μ -integrabel, wobei μ ein endliches Maß auf $\mathcal{F}_1$ sei. Dann gibt es eine Folge von Mengenfamilien

$$(\{E_i^{(n)}\}_{i=1}^{k_n})_{n \in \mathbb{N}}, \quad E_i^{(n)} \in \mathcal{F}_0$$

und eine Folge von Funktionen

$$\begin{aligned} h_n : X &\rightarrow \mathbb{R}^+ \cup \{0\} \\ h_n &= \sum_{i=1}^{k_n} \alpha_i^{(n)} \chi_{E_i^{(n)}} \end{aligned} \quad (2.31)$$

so daß für alle $x \in X$:

$$\lim_{n \rightarrow \infty} h_n(x) = h(x) \quad [m] \quad (2.32)$$

$\lim_{n \rightarrow \infty} h_n$ ist meßbar bezüglich der kleinsten σ -Algebra

$$\sigma(\{E_i^{(n)} \mid 1 \leq i \leq k_n, n \in \mathbb{N}\})$$

die alle $E_i^{(n)}$ umfaßt.

Es folgt der Hauptsatz:

SATZ 2.9

$\mathcal{X}$ sei separabel. Dann gilt für alle $t \in [0,1]$:

$$I(P, Q)(t) = \inf_{\mathcal{P}} \sum_{i=1}^m P(E_i)^t Q(E_i)^{1-t} \quad (2.33)$$

$\mathcal{P} = \{E_i \mid 1 \leq i \leq m\}$
 $\mathcal{P}$ endliche Partition von X

Beweis:

Ohne Einschränkung sei $\lambda = \frac{P+Q}{2}$, d.h. selbst ein Wahrscheinlichkeitsmaß (siehe auch S. 6). $\mathcal{P}_0 \subset \mathcal{P}$ sei wie in Lemma 2.8 gewählt. Aufgrund dieses Lemmas gibt es zu f und g Funktionen ($f = \frac{dP}{d\lambda}$, $g = \frac{dQ}{d\lambda}$)

$$f_n, g_n : X \rightarrow \mathbb{R}^+ \cup \{0\}$$

sowie Mengenfamilien

$$\{E_i^{(n)}\}_{1 \leq i \leq k_n}, \quad \{E_i'^{(n)}\}_{1 \leq i \leq k_n'}$$

aus $\mathcal{P}_0$ mit

$$f_n = \sum_{i=1}^{k_n} \beta_i^{(n)} \chi_{E_i^{(n)}} \quad (2.34)$$

$$g_n = \sum_{i=1}^{k_n'} \beta_i'^{(n)} \chi_{E_i'^{(n)}} \quad (2.35)$$

$\beta_i^{(n)}, \beta_i'^{(n)} \in \mathbb{R}^+$ und

$$\lim_{n \rightarrow \infty} f_n = f \quad [\lambda] \quad (2.36)$$

$$\lim_{n \rightarrow \infty} g_n = g \quad [\lambda] \quad (2.37)$$

Aus Stetigkeitsgründen gilt dann für $t \in [0, 1]$:

$$\lim_{n \rightarrow \infty} f_n^t = f^t \quad [\lambda] \quad (2.38)$$

$$\lim_{n \rightarrow \infty} g_n^{1-t} = g^{1-t} \quad [\lambda] \quad (2.39)$$

$$\lim_{n \rightarrow \infty} f_n^t g_n^{1-t} = f^t g^{1-t} \quad [\lambda] \quad (2.40)$$

Damit können f , g und $f^t g^{1-t}$ als meßbar bez. der kleinsten σ -Algebra $\mathcal{F}$ gewählt werden, die alle

$$\{E_i^{(n)}\}_{1 \leq i \leq k_n}, \quad \{E_i'^{(n)}\}_{1 \leq i \leq k'_n}$$

umfaßt, d.h.

$$= \sigma(\{E_i^{(n)} \mid 1 \leq i \leq k_n, n \in \mathbb{N}\} \cup \{E_i'^{(n)} \mid 1 \leq i \leq k'_n, n \in \mathbb{N}\}) \quad (2.41)$$

Man erhält dann die folgende Relation:

$$\mathcal{F} \subset \sigma(\mathcal{F}_0) \subset \mathcal{X} \quad (2.42)$$

Induktiv wird jetzt eine Folge $\{\mathcal{T}_n\}_{n \in \mathbb{N}}$ von Partitionen des Raumes X konstruiert, indem man sich eines Lemmas bedient, welches im Buch von Kingman-Taylor [36] S. 11 zu finden ist.

Es lautet:

Ist $\{M_i\}_{i \in I}$ eine endliche oder abzählbar unendliche Familie von Mengen, so gibt es Mengen $N_i \subset M_i$, wobei die N_i paarweise disjunkt sind, so daß

$$\bigcup_{i \in I} M_i = \bigcup_{i \in I} N_i \quad (2.43)$$

Mit anderen Worten: Jede höchstens abzählbare Vereinigung von Mengen läßt sich als disjunkte Vereinigung von Teilmengen auffassen.

Nun zurück zur Konstruktion.

Sei $n = 1$. Man betrachte die Mengen $E_i^{(1)}$, $1 \leq i \leq k_1$ sowie $E_i'^{(1)}$, $1 \leq i \leq k'_1$.

Zunächst können wir

$$\bigcup_{i=1}^{k_1} E_i^{(1)} \cup \bigcup_{i=1}^{k'_1} E_i'^{(1)}$$

im Sinne des vorigen Lemmas disjunkt zerlegen in Teilmengen von $E_i^{(1)}$ bzw. $E_i'^{(1)}$. Auf diese Weise erhält man $(k_1 + k'_1)$ disjunkte Teilmengen $T_j^{(1)}$ mit

$$\bigcup_{i=1}^{k_1} E_i^{(1)} \cup \bigcup_{i=1}^{k'_1} E_i'^{(1)} = \bigcup_{j=1}^{k_1 + k'_1} T_j^{(1)} \quad (2.44)$$

Setzt man

$$T_{k_1 + k'_1 + 1} : = X \setminus \bigcup_{j=1}^{k_1 + k'_1} T_j^{(1)} \quad (2.45)$$

so ist

$$\mathcal{T}_1 : = \{T_j^{(1)} \mid 1 \leq j \leq k_1 + k'_1 + 1\} \quad (2.46)$$

eine endliche Partition von X .

Seien jetzt $\mathcal{T}_1, \dots, \mathcal{T}_n$ bereits konstruiert. Man betrachte jetzt für $n+1$ die Mengen

$$E_i^{(n+1)}, E_i'^{(n+1)} \quad 1 \leq i \leq k_{n+1}, \quad 1 \leq i \leq k'_{n+1}$$

Diese Mengen liefern analog zum Fall $n=1$ eine endliche Partition von X , die mit $\mathcal{T}'$ bezeichnet werde. Man setze dann

$$\mathcal{T}_{n+1} : = \{T_j^{(n)} \cap T' \mid T_j^{(n)} \in \mathcal{T}_n, T' \in \mathcal{T}'\} \quad (2.47)$$

Damit ist $\mathcal{T}_{n+1}$ ebenfalls eine endliche Partition von X . $\mathcal{F}_n$ sei die endliche, durch $\mathcal{T}_n$ erzeugte σ -Algebra, d.h. für $n \in \mathbb{N}$ ist

$$\mathcal{F}_n : = \sigma(\mathcal{T}_n)$$

Aufgrund obiger Konstruktion gilt:

$$(i) \quad \mathcal{T}_1 \subset \mathcal{T}_2 \subset \dots \quad (2.48)$$

$$(ii) \quad \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \quad (2.49)$$

$$(iii) \quad \mathcal{F}_n \supset \sigma(\{E_i^{(m)} \mid 1 \leq i \leq k_m, 1 \leq m \leq n\} \cup \{E_i'^{(m)} \mid 1 \leq i \leq k'_m, 1 \leq m \leq n\}) \quad (2.50)$$

$$(iv) \quad \mathcal{F}_\infty := \sigma \left(\bigcup_{n=1}^{\infty} \mathcal{F}_n \right) \supset \mathcal{F} \quad (2.51)$$

Wegen (iv) sind f und g auch meßbar bez. $\mathcal{F}_\infty$. Auf X definiere man nun die reellen Funktionen

$$\phi_n(x) = \begin{cases} \frac{P(T_j^{(n)})}{\lambda(T_j^{(n)})} & \text{falls } x \in T_j^{(n)}, \lambda(T_j^{(n)}) > 0 \\ 0 & \text{falls } x \in T_j^{(n)}, \lambda(T_j^{(n)}) = 0 \end{cases} \quad (2.52)$$

$$\psi_n(x) = \begin{cases} \frac{Q(T_j^{(n)})}{\lambda(T_j^{(n)})} & \text{falls } x \in T_j^{(n)}, \lambda(T_j^{(n)}) > 0 \\ 0 & \text{falls } x \in T_j^{(n)}, \lambda(T_j^{(n)}) = 0 \end{cases} \quad (2.53)$$

Wie man leicht sieht, bilden

$$\{\phi_n, \mathcal{F}_n\}_{n \in \mathbb{N}}, \quad \{\psi_n, \mathcal{F}_n\}_{n \in \mathbb{N}}$$

jeweils ein Martingal. (Vgl. Doob [11], § 8, S. 343, 344.)

Wie man weiter sieht, gilt:

$$E_\lambda(f | \mathcal{F}_n) = \phi_n \quad [\lambda] \quad (2.54)$$

$$E_\lambda(g | \mathcal{F}_n) = \psi_n \quad [\lambda] \quad (2.55)$$

Wegen Doob [11] Theorem 4.3, S. 331 ist aber

$$\lim_{n \rightarrow \infty} E_\lambda(f | \mathcal{F}_n) = E_\lambda(f | \mathcal{F}_\infty) \quad [\lambda] \quad (2.56)$$

$$\lim_{n \rightarrow \infty} E_\lambda(g | \mathcal{F}_n) = E_\lambda(g | \mathcal{F}_\infty) \quad [\lambda] \quad (2.57)$$

Damit gilt, da f und g meßbar sind bezüglich $\mathcal{F}_\infty$:

$$\lim_{n \rightarrow \infty} \phi_n = f \quad [\lambda] \quad (2.58)$$

$$\lim_{n \rightarrow \infty} \psi_n = g \quad [\lambda] \quad (2.59)$$

Also auch

$$\lim_{n \rightarrow \infty} \phi_n^t \psi_n^{1-t} = f^t g^{1-t} \quad [\lambda] \quad (2.60)$$

Nun ist aber für $x \in X$

$$\phi_n^t \psi_n^{1-t}(x) = \begin{cases} \frac{P(T_j^{(n)})^t Q(T_j^{(n)})^{1-t}}{\lambda(T_j^{(n)})} & \text{falls } x \in T_j^{(n)}, \lambda(T_j^{(n)}) > 0 \\ 0 & \text{falls } x \in T_j^{(n)}, \lambda(T_j^{(n)}) = 0 \end{cases} \quad (2.61)$$

Andererseits gilt

$$P(T_j^{(n)})^t Q(T_j^{(n)})^{1-t} \leq tP(T_j^{(n)}) + (1-t) Q(T_j^{(n)}) \quad (2.62)$$

$$\leq P(T_j^{(n)}) + Q(T_j^{(n)})$$

$$= 2\lambda(T_j^{(n)}) \quad (2.63)$$

und deshalb

$$|\phi_n^t \psi_n^{1-t}| \leq 2$$

Wegen Halmos [20], Theorem D, S. 110 (Lebesgue's bounded convergence Theorem) ist

$$\lim_{n \rightarrow \infty} \int |\phi_n^t \psi_n^{1-t} - f^t g^{1-t}| d\lambda = 0 \quad (2.64)$$

Dies impliziert

$$\lim_{n \rightarrow \infty} \left| \int \phi_n^t \psi_n^{1-t} d\lambda - \int f^t g^{1-t} d\lambda \right| = 0 \quad (2.65)$$

Nun ist aber

$$\int_X \phi_n^t \psi_n^{1-t} = \sum_j \int_{T_j^{(n)}} \phi_n^t \psi_n^{1-t} \quad (2.66)$$

$$= \sum_j \frac{P(T_j^{(n)})^t Q(T_j^{(n)})^{1-t}}{\lambda(T_j^{(n)})} \lambda(T_j^{(n)}) \quad (2.67)$$

$$= \sum_j P(T_j^{(n)})^t Q(T_j^{(n)})^{1-t} \quad (2.68)$$

wobei zunächst eigentlich nur über die Indizes summiert wurde, für die $\lambda(T_j^{(n)}) > 0$ ist. Die Gleichung (2.68) bleibt aber für die restlichen Indizes unverändert, da für alle j mit $\lambda(T_j^{(n)}) = 0$ stets auch $P(T_j^{(n)}) = 0$ und $Q(T_j^{(n)}) = 0$ folgt (λ dominiert P und Q).

Setzt man in (2.65) ein, so resultiert

$$\lim_{n \rightarrow \infty} \left| \sum_j P(T_j^{(n)})^t Q(T_j^{(n)})^{1-t} - \int f^t g^{1-t} d\lambda \right| = 0 \quad (2.69)$$

oder aber

$$\lim_{n \rightarrow \infty} \left| \sum_j P(T_j^{(n)})^t Q(T_j^{(n)})^{1-t} - I(P, Q)(t) \right| = 0 \quad (2.70)$$

Dies genügt aber zu zeigen.

Als jetzt einfache Folgerung erhält man

SATZ 2.10

$\mathcal{X}$ sei separabel. Dann gilt für alle $t \in [0, 1]$:

$$I(P, Q)(t) = \inf_{\begin{array}{l} \mathcal{T} = \{E_i \mid i \in \mathbb{N}\} \\ \mathcal{T} \text{ ist abzählbar} \\ \text{unendliche Partition von } X \end{array}} \sum_{i=1}^{\infty} P(E_i)^t Q(E_i)^{1-t} \quad (2.71)$$

Beweis:

" $\leq$ " klar wegen (1.76).

" $\geq$ ". Seien $\mathcal{T}'$ beliebige abzählbar unendliche Partition von X und $\mathcal{T}''$ beliebige endliche Partition von X . Dann ist

$$\mathcal{T}' \wedge \mathcal{T}'' := \{T' \cap T'' \mid T' \in \mathcal{T}', T'' \in \mathcal{T}''\} \quad (2.72)$$

eine abzählbar unendliche Partition mit $\mathcal{T}'' < \mathcal{T}' \wedge \mathcal{T}''$.

Es ist

$$\inf_{\mathcal{I} = \{E_i | i \in \mathbb{N}\}} \sum_{i=1}^{\infty} P(E_i)^t Q(E_i)^{1-t} \leq \sum_{E' \in \mathcal{I}' \wedge \mathcal{I}''} P(E')^t Q(E')^{1-t} \quad (2.73)$$

$\mathcal{I}$ ist abzählbar
unendliche Partition von X

Da aber auch wegen Lemma 2.1

$$\sum_{E' \in \mathcal{I}' \wedge \mathcal{I}''} P(E')^t Q(E')^{1-t} \leq \sum_{E'' \in \mathcal{I}''} P(E'')^t Q(E'')^{1-t} \quad (2.74)$$

so folgt wegen Satz 2.9

$$I(P, Q)(t) \geq \inf_{\mathcal{I} = \{E_i | i \in \mathbb{N}\}} \sum_{i=1}^{\infty} P(E_i)^t Q(E_i)^{1-t} \quad (2.75)$$

$\mathcal{I}$ ist abzählbar
unendliche Partition von X

III. Kapitel

Gestutzte Verteilungen

Bekanntlich gilt für eine normalverteilte Zufallsvariable Z mit $E(Z) = \mu$ und $\text{Var}(Z) = \sigma^2$, daß

$$\text{Prob}(|Z - \mu| > 3\sigma) < 0.003 \quad (3.1)$$

Fast die gesamte "Wahrscheinlichkeitsmasse" ist in dem 3-fachen zentralen Schwankungsintervall $[\mu - 3\sigma, \mu + 3\sigma]$ konzentriert. Man kann dieser Tatsache Rechnung tragen, daß man die Dichte f von Z so "zurechtstutzt", daß f außerhalb des Intervalls $[\mu - 3\sigma, \mu + 3\sigma]$ verschwindet, d.h. den ursprünglichen Träger von f , das Intervall $(-\infty, \infty)$ auf ein kleineres, aber relevantes Intervall zusammendrückt. Wie die AEF auf diese "Stutzung" reagiert, wird im folgenden behandelt.

Zunächst jedoch wird diese neue Begriffsbildung exakt definiert. (Siehe z.B. auch Fisz [15] S. 72)

Sei $\phi : X \rightarrow \mathbb{R}$ $(\mathcal{X} - \mathcal{C}_0^1)$ -meßbar mit

$$0 < \int \phi f \, d\lambda < \infty \quad (3.2)$$

$$0 < \int \phi g \, d\lambda < \infty \quad (3.3)$$

wobei f und g wie oben gewählt seien.

Dann liefert ϕ die gestutzten Wahrscheinlichkeitsmaße P^ϕ und Q^ϕ auf $\mathcal{X}$ definiert durch

$$P^\phi(E) = \frac{\int_E \phi f \, d\lambda}{\int \phi f \, d\lambda} \quad \text{für } E \in \mathcal{X} \quad (3.4)$$

$$Q^\phi(E) = \frac{\int_E \phi g \, d\lambda}{\int \phi g \, d\lambda} \quad \text{für } E \in \mathcal{X} \quad (3.5)$$

Dann kann man auch so schreiben

$$\frac{dP^\phi}{d\lambda} = \frac{1}{\int \phi f \, d\lambda} \phi f \quad (3.6)$$

$$\frac{dQ^\phi}{d\lambda} = \frac{1}{\int \phi g \, d\lambda} \phi g \quad (3.7)$$

LEMMA 3.1

Seien P^ϕ und Q^ϕ die gestutzten Maße bez. P und Q .
Dann gilt für alle $t \in [0,1]$

$$(i) \quad I(P^\phi, Q^\phi)(t) = \left[\frac{1}{\int \phi f \, d\lambda} \right]^t \left[\frac{1}{\int \phi g \, d\lambda} \right]^{1-t} \int \phi f^t g^{1-t} \, d\lambda \quad (3.8)$$

(ii) Ist $\phi = \chi_E$ mit $E \in \mathcal{X}$, so gilt für alle $t \in [0,1]$:

$$I(P^\phi, Q^\phi)(t) = \left[\frac{1}{P(E)} \right]^t \left[\frac{1}{Q(E)} \right]^{1-t} I_E(P, Q)(t) \quad (3.9)$$

Beweis:

Folgt direkt aus den Definitionen.

Wir geben jetzt einen interessanten Zusammenhang zwischen AEF, gestutzter Verteilung und meßbarer Transformation an.

SATZ 3.2

$(Y, \mathcal{V})$ sei ein Meßraum und

$$T : X \rightarrow Y$$

sei $(\mathcal{X} - \mathcal{V})$ - meßbar.

Weiterhin sei $F \in \mathcal{V}$ mit $T(P)(F) > 0$, $T(Q)(F) > 0$.

Man setze $\phi := \chi_{T^{-1}(F)}$. Dann gilt für alle $t \in [0,1]$:

$$(i) \quad I(P^\phi, Q^\phi)(t) = \left[\frac{1}{T(P)(F)} \right]^t \left[\frac{1}{T(Q)(F)} \right]^{1-t} I_{T^{-1}(F)}(P, Q)(t) \quad (3.10)$$

$$(ii) \quad I(T(P^\phi), T(Q^\phi))(t) = \left[\frac{1}{T(P)(F)} \right]^t \left[\frac{1}{T(Q)(F)} \right]^{1-t} I_F(T(P), T(Q))(t) \quad (3.11)$$

$$(iii) \quad \frac{I_F(T(P), T(Q))(t)}{I_{T^{-1}(F)}(P, Q)(t)} = \frac{I(T(P^\phi), T(Q^\phi))(t)}{I(P^\phi, Q^\phi)(t)} \quad (3.12)$$

Beweis:

(i) Folgt sofort aus Lemma 3.1.

(ii) Nach Definition ist

$$I(T(P^\phi), T(Q^\phi))(t) = \int E_\lambda \left(\frac{dP^\phi}{d\lambda} \mid T=y \right)^t E_\lambda \left(\frac{dQ^\phi}{d\lambda} \mid T=y \right) dT(\lambda)(y) \quad (3.13)$$

wobei

$$\frac{dP^\phi}{d\lambda} = \frac{1}{P(T^{-1}(F))} \cdot f \cdot X_{T^{-1}(F)} \quad (3.14)$$

$$\frac{dQ^\phi}{d\lambda} = \frac{1}{Q(T^{-1}(F))} \cdot g \cdot X_{T^{-1}(F)} \quad (3.15)$$

Weiterhin ist, wie man leicht sieht

$$X_{T^{-1}(F)} = X_F \circ T$$

Daraus folgt (siehe Halmos [19]):

$$E_\lambda \left(\frac{dP^\phi}{d\lambda} \mid T=y \right) = \frac{1}{P(T^{-1}(F))} \cdot X_F \cdot E_\lambda(f \mid T=y) \quad [T(\lambda)] \quad (3.16)$$

$$E_\lambda \left(\frac{dQ^\phi}{d\lambda} \mid T=y \right) = \frac{1}{Q(T^{-1}(F))} \cdot X_F \cdot E_\lambda(g \mid T=y) \quad [T(\lambda)] \quad (3.17)$$

und weiter aus (3.13)

$$I(T(P^\phi), T(Q^\phi))(t) = \left[\frac{1}{T(P)(F)} \right]^t \left[\frac{1}{T(Q)(F)} \right]^{1-t} \int_F E_\lambda(f \mid T=y)^t E_\lambda(g \mid T=y)^{1-t} dT(\lambda)(y) \quad (3.18)$$

Damit ist (ii) gezeigt.

(iii) Folgt direkt aus (i) und (ii).

Bemerkung:

Setzt man $F = Y$, so ist

$$\frac{I(T(P), T(Q))(t)}{I(P, Q)(t)} = \frac{I(T(P^\phi), T(Q^\phi))(t)}{I(P^\phi, Q^\phi)(t)} \quad (3.19)$$

Dies ist aber eine Trivialität, da dann

$$P^\phi = P, \quad Q^\phi = Q$$

Der folgende Satz ist ein Analogon zu einem Ergebnis von Kullback [40]

Satz 2, Corollary 3.1:

SATZ 3.3

(i) Sei $E \in \mathcal{X}$, und es gelte

$$\frac{f}{g}(x) = \frac{P(E)}{Q(E)} \quad [\lambda] \quad \text{für } x \in E \quad (3.20)$$

dann ist für alle $t \in [0,1]$

$$I_E(P,Q)(t) = P(E)^t Q(E)^{1-t} \quad (3.21)$$

(ii) Sei $E \in \mathcal{X}$ und gelte für ein $t_0 \in (0,1)$

$$I_E(P,Q)(t_0) = P(E)^{t_0} Q(E)^{1-t_0} \quad (3.22)$$

so ist

$$\frac{f}{g}(x) = \frac{P(E)}{Q(E)} \quad [\lambda] \quad \text{für } x \in E \quad (3.23)$$

Beweis:

(i) folgt unmittelbar aus der Kette von Gleichungen

$$\begin{aligned} I_E(P,Q)(t) &= \int_E f^t g^{1-t} d\lambda \\ &= \int_E f^t g^{-t} g d\lambda \\ &= \int_E \left[\frac{P(E)}{Q(E)} \right]^t g d\lambda \\ &= \left[\frac{P(E)}{Q(E)} \right]^t Q(E) \\ &= P(E)^t Q(E)^{1-t} \end{aligned} \quad (3.24)$$

(ii) Gilt $I_E(P,Q)(t_0) = P(E)^{t_0} Q(E)^{1-t_0}$ für ein $t_0 \in (0,1)$, so ist wegen Lemma 3.1, (3.9):

$$I(P^\phi, Q^\phi)(t_0) = 1 \quad (3.25)$$

mit $\phi = \chi_E$.

Da

$$I(P^\phi, Q^\phi)(\cdot) : [0,1] \rightarrow [0,1]$$

eine konvexe Funktion ist, folgt aus Lemma 1.2:

$$I(P^\phi, Q^\phi)(t) = 1 \quad (3.26)$$

für alle $t \in [0,1]$.

Dann ist auch $I(P^\phi, Q^\phi)(\frac{1}{2}) = 1$ und damit $P^\phi = Q^\phi$ wegen [1].

Dann folgt jedoch direkt

$$\frac{x_E f}{P(E)} = \frac{x_E g}{Q(E)} \quad [\lambda] \quad (3.27)$$

d.h. die Behauptung

$$\frac{f}{g}(x) = \frac{P(E)}{Q(E)} \quad \text{für } x \in E \quad [\lambda] \quad (3.28)$$

ist erfüllt.

Es schließt sich an:

KOROLLAR 3.4

$\{E_i\}_{i \in I}$ sei endliche oder abzählbar unendliche Partition von X . Dann gilt:

(i) Ist

$$\frac{f}{g}(x) = \frac{P(E_i)}{Q(E_i)} \quad x \in E_i \quad [\lambda] \quad i \in I \quad (3.29)$$

so folgt für alle $t \in [0,1]$:

$$I(P,Q)(t) = \sum_{i \in I} P(E_i)^t Q(E_i)^{1-t} \quad (3.30)$$

(ii) Ist

$$I(P,Q)(t_0) = \sum_{i \in I} P(E_i)^{t_0} Q(E_i)^{1-t_0} \quad (3.31)$$

für ein $t_0 \in (0,1)$, so folgt:

$$\frac{f}{g}(x) = \frac{P(E_i)}{Q(E_i)} \quad x \in E_i \quad [\lambda] \quad i \in I \quad (3.32)$$

Beweis:

(i) Folgt direkt aus Satz 3.3 und der Beziehung

$$I(P,Q)(t) = \sum_{i \in I} \int_{E_i} f^t g^{1-t} d\lambda \quad (3.33)$$

(ii) Angenommen, es gäbe ein $j \in I$ mit $\lambda(E_j) > 0$, so daß

$$\frac{f}{g}(x) \neq \frac{P(E_j)}{Q(E_j)} \quad (3.34)$$

für mindestens ein $x \in E_j$.

Wegen Satz 3.3 ist dann

$$I_{E_j}(P, Q)(t_0) < P(E_j)^{t_0} Q(E_j)^{1-t_0} \quad (3.35)$$

Da stets

$$I_{E_i}(P, Q)(t_0) \leq P(E_i)^{t_0} Q(E_i)^{1-t_0} \quad (3.36)$$

so folgt

$$\sum_{i \in I} I_{E_i}(P, Q)(t_0) < \sum_{i \in I} P(E_i)^{t_0} Q(E_i)^{1-t_0}$$

Das ist jedoch ein Widerspruch zur Voraussetzung.

Es folgen einige illustrierende Beispiele.

Beispiel 1:

P, Q mögen folgende Dichte haben (bez. dem Borel-Maß auf $\mathcal{L}^1$):

$$f(x) = \begin{cases} \frac{1}{\theta_1} & \text{falls } x \in [0, \theta_1] \\ 0 & \text{sonst} \end{cases} \quad (3.37)$$

$$g(x) = \begin{cases} \frac{1}{\theta_2} & \text{falls } x \in [0, \theta_2] \\ 0 & \text{sonst} \end{cases} \quad (3.38)$$

mit $0 < \theta_1 < \theta_2$.

$\{E_i\}_{i \in I}$ sei abzählbar unendliche oder endliche Partition von $X = \mathbb{R}$, bestehend aus halboffenen Intervallen der Form $[a_i, b_i)$, wobei der Fall $a_i = -\infty$ bzw. $b_i = \infty$ zugelassen sei. Dann gilt:

$$I(P, Q)(t) = \sum_{i \in I} P(E_i)^t Q(E_i)^{1-t} \quad (3.39)$$

Ist nämlich $[a_i, b_i) \subset \text{TR}(f)$, so gilt

$$P([a_i, b_i]) = \frac{b_i - a_i}{\theta_1} \quad (3.40)$$

$$Q([a_i, b_i]) = \frac{b_i - a_i}{\theta_2} \quad (3.41)$$

Damit ist

$$\frac{P([a_i, b_i])}{Q([a_i, b_i])} = \frac{\theta_2}{\theta_1} \quad (3.42)$$

Andererseits gilt für $x \in [a_i, b_i]$:

$$\frac{f(x)}{g(x)} = \frac{1/\theta_1}{1/\theta_2} = \frac{\theta_2}{\theta_1} \quad (3.43)$$

Aus Korollar 3.4 folgt dann direkt die Beziehung

$$I(P, Q)(t) = \sum_{i \in I} P(E_i)^t Q(E_i)^{1-t} \quad (3.44)$$

mit $E_i = [a_i, b_i]$.

(3.44) kann man übrigens auch direkt beweisen:

Ohne Einschränkung kann man sich (wegen der Kompaktheit von $TR(f)$) auf endlich viele $[a_i, b_i] \subset TR(f)$ beschränken, d.h.

$$TR(f) = \bigcup_{i=1}^n [a_i, b_i] \quad (3.45)$$

wobei man annehmen kann, daß $a_1 = 0, b_n = \theta_1$.

Dann ist

$$\begin{aligned} \sum_{i \in I} P(E_i)^t Q(E_i)^{1-t} &= \sum_{i \in I} P([a_i, b_i])^t Q([a_i, b_i])^{1-t} \\ &= \sum_{i=1}^n \left[\frac{b_i - a_i}{\theta_1} \right]^t \left[\frac{b_i - a_i}{\theta_2} \right]^{1-t} \\ &= \sum_{i=1}^n (b_i - a_i) \theta_1^{-t} \theta_2^{t-1} \\ &= \theta_1 \theta_1^{-t} \theta_2^{t-1} \\ &= \left(\frac{\theta_1}{\theta_2} \right)^{1-t} \end{aligned} \quad (3.46)$$

Dem Anhang entnimmt man, daß

$$I(P,Q)(t) = \left(\frac{\theta_1}{\theta_2} \right)^{1-t} \quad (3.47)$$

Zum folgenden Beispiel siehe man auch Kullback [40] S. 17, Bsp. 3.2.

Beispiel 2:

Sei $X = \mathbb{N}$, $\mathcal{X} = \mathcal{P}(\mathbb{N})$. P, Q seien diskrete Wahrscheinlichkeitsmaße auf $\mathcal{X}$. Weiterhin sei $E \subset \mathbb{N}$, und es gelte

$$\begin{aligned} P(\{x_i\}) &= : p_i > 0 \\ Q(\{x_i\}) &= : q_i > 0 \end{aligned} \quad (3.48)$$

für alle $x_i \in E$.

Es ist für $t \in [0,1]$:

$$I_E(P,Q)(t) = \sum_{x_i \in E} p_i^t q_i^{1-t} \quad (3.49)$$

$$P(E)^t Q(E)^{1-t} = \left(\sum_{x_i \in E} p_i \right)^t \left(\sum_{x_i \in E} q_i \right)^{1-t} \quad (3.50)$$

Wenn für alle $x_i \in E$

$$\frac{p_i}{q_i} = \frac{\sum_{x_j \in E} p_j}{\sum_{x_j \in E} q_j} \quad (3.51)$$

gilt, so hat man wegen Satz 3.3:

$$I_E(P,Q)(t) = P(E)^t Q(E)^{1-t} \quad (3.52)$$

d.h.

$$\sum_{x_i \in E} p_i^t q_i^{1-t} = \left(\sum_{x_i \in E} p_i \right)^t \left(\sum_{x_i \in E} q_i \right)^{1-t} \quad (3.53)$$

Gilt umgekehrt für ein $t_0 \in (0,1)$

$$\sum_{x_i \in E} p_i^{t_0} q_i^{1-t_0} = \left(\sum_{x_i \in E} p_i \right)^{t_0} \left(\sum_{x_i \in E} q_i \right)^{1-t_0} \quad (3.54)$$

so gilt für alle $x_i \in E$

$$\frac{p_i}{q_i} = \frac{\sum_{x_j \in E} p_j}{\sum_{x_j \in E} q_j} \quad (3.55)$$

Zum nachstehenden Beispiel siehe man wiederum Kullback [40] Bsp. 3.3, S. 17.

Beispiel 3:

Sei $X = \mathbb{N} \cup \{0\}$, $\mathcal{X} = \mathcal{P}(X)$, $E_1 := \{0\}$, $E_2 := \{1\}$, $E_3 := \{2, 3, \dots\}$

P, Q seien Wahrscheinlichkeitsmaße mit Poissondichte (vgl. Anhang), d.h.

$$P(x) = e^{-\lambda_1} \frac{\lambda_1^x}{x!} \quad (3.56)$$

$$Q(x) = e^{-\lambda_2} \frac{\lambda_2^x}{x!} \quad (3.57)$$

mit $\lambda_1, \lambda_2 > 0$, $x \in \mathbb{N} \cup \{0\}$

$\{E_i\}_{i=1}^3$ bildet eine Partition von X . Es ist

$$\begin{aligned} P(E_1) &= e^{-\lambda_1} & Q(E_1) &= e^{-\lambda_2} \\ P(E_2) &= \lambda_1 e^{-\lambda_1} & Q(E_2) &= \lambda_2 e^{-\lambda_2} \\ P(E_3) &= 1 - (1 + \lambda_1) e^{-\lambda_1} & Q(E_3) &= 1 - (1 + \lambda_2) e^{-\lambda_2} \end{aligned} \quad (3.58)$$

und somit

$$\begin{aligned} \sum_{i=1}^3 P(E_i)^t Q(E_i)^{1-t} &= e^{-\lambda_1 t - \lambda_2 (1-t)} (1 + \lambda_1^t \lambda_2^{1-t}) \\ &\quad + (1 - e^{-\lambda_1} (1 + \lambda_1))^t (1 - e^{-\lambda_2} (1 + \lambda_2))^{1-t} \end{aligned} \quad (3.59)$$

Dem Anhang entnimmt man, daß

$$I(P, Q)(t) = e^{-\lambda_1 t - \lambda_2 (1-t) + \lambda_1^t \lambda_2^{1-t}} \quad (3.60)$$

Die Gleichheit mit (3.60) tritt ein, wenn

$$\frac{P(x)}{Q(x)} = \frac{P(E_i)}{Q(E_i)} \quad \text{für alle } x \in E_i \quad (3.61)$$

(i = 1, 2, 3)

Da dies für i = 1, 2 ohnehin erfüllt ist, genügt als hinreichende Bedingung für die Gleichheit

$$\frac{e^{-\lambda_1} \lambda_1^x}{e^{-\lambda_2} \lambda_2^x} = \frac{1 - e^{-\lambda_1}(1 + \lambda_1)}{1 - e^{-\lambda_2}(1 + \lambda_2)} \quad x = 2, 3, \dots \quad (3.62)$$

Gilt umgekehrt

$$I(P, Q)(t) = \sum_{i=1}^3 P(E_i)^{t_0} Q(E_i)^{1-t_0} \quad (3.63)$$

für ein $t_0 \in (0, 1)$, so ist

$$\frac{e^{-\lambda_1} \lambda_1^x}{e^{-\lambda_2} \lambda_2^x} = \frac{1 - e^{-\lambda_1}(1 + \lambda_1)}{1 - e^{-\lambda_2}(1 + \lambda_2)} \quad (3.64)$$

für alle $x = 2, 3, \dots$

IV. Kapitel

Suffizienz

Die suffizienten Statistiken sind ein wichtiges Hilfsmittel, Stichprobenwerte so zu transformieren, daß das jeweils anstehende Problem (Schätzproblem, Testproblem) sich leichter lösen läßt, aber dabei noch alle Information, die relevant ist, erhalten bleibt. Die Definition der Suffizienz soll hier nicht gebracht werden; sie findet sich in fast allen Lehrbüchern der Statistik, besonders anschaulich in Witting [56] S. 136.

Die Entscheidung, ob eine vorliegende Statistik suffizient ist, geschieht zumeist durch das allgemeine Faktorisierungstheorem (siehe Halmos [19]). Es sei hier noch ein weiteres Entscheidungskriterium hinzugefügt.

SATZ 4.1

$(Y, \mathcal{W})$ sei ein Maßraum

$$T : X \rightarrow Y$$

sei $(\mathcal{X}, \mathcal{W})$ - meßbar. Außerdem seien $I(P, Q)(\cdot)$ und $I(T(P), T(Q))(\cdot)$ in einer offenen Umgebung von $t_0 = 1$ differenzierbar, und es gelte dort

$$\frac{d}{dt} [I(P, Q)(t)] = \int f^t g^{1-t} \log \frac{f}{g} d\lambda \quad (4.1)$$

$$\frac{d}{dt} [I(T(P), T(Q))(t)] = \int E_\lambda(f|T=y)^t E_\lambda(g|T=y)^{1-t} \log \frac{E_\lambda(f|T=y)}{E_\lambda(g|T=y)} dT(\lambda)(y) \quad (4.2)$$

und

$$I(P, Q)(t) = I(T(P), T(Q))(t) \quad (4.3)$$

Dann gilt: T ist suffizient für $\{P, Q\}$

Beweis:

Sei U die offene Umgebung von $t_0 = 1$ aus der Voraussetzung. Dann ist für alle $t \in U \setminus \{t_0\}$

$$I_t(P||Q) = I_t(T(P)||T(Q)) \quad (4.4)$$

Hierzu brachte man (1.23) auf S. 8. Damit existieren und gleichen sich die nachstehenden Grenzwerte

$$\lim_{t \rightarrow 1} I_t(P||Q) \quad \text{und} \quad (4.5)$$

$$\lim_{t \rightarrow 1} I_t(T(P)||T(Q))$$

d.h.

$$K(P, Q) = K(T(P), T(Q)) \quad (4.6)$$

Siehe dazu S. 16, (1.70).

Wegen Kullback [40], Theorem 4.2, S. 21 (vgl. auch die Bemerkung auf S. 21) gilt dann die Behauptung.

Interessant ist die Umkehrung, wobei man sogar mit schwächeren Voraussetzungen auskommt.

SATZ 4.2

$(Y, \mathcal{M})$ sei ein Maßraum.

$$T : X \rightarrow Y$$

sei $(\mathcal{K} - \mathcal{M})$ - meßbar und suffizient für $\{P, Q\}$. Dann gilt für alle $t \in [0, 1]$:

$$I(P, Q)(t) = I(T(P), T(Q))(t) \quad (4.7)$$

Beweis:

Man wähle $\lambda = \frac{P + Q}{2}$. Wegen Witting [56], S. 149 gibt es dann

$(\mathcal{M} - \mathcal{B}^1)$ - meßbare Funktionen

$$h_P, h_Q : Y \rightarrow \mathbb{R}$$

mit

$$f = h_P \circ T \quad [\lambda] \quad (4.8)$$

$$g = h_Q \circ T \quad [\lambda] \quad (4.9)$$

Wegen Halmos-Savage [19], S. 231 gilt dann

$$E_\lambda(f|T=y) = h_P(y) \quad [T(\lambda)] \quad (4.10)$$

$$E_\lambda(g|T=y) = h_Q(y) \quad [T(\lambda)] \quad (4.11)$$

Nun ist aber für $t \in [0, 1]$

$$I(T(P), T(Q))(t) = \int E_\lambda(f|T=y)^t E_\lambda(g|T=y)^{1-t} dT(\lambda)(y) \quad (4.12)$$

Aus (4.10) bzw. (4.11) erhält man dann

$$\begin{aligned}
 I(T(P), T(Q))(t) &= \int h_P(y)^t h_Q(y)^{1-t} dT(\lambda)(y) \\
 &= \int h_P(T(x))^t h_Q(T(x))^{1-t} d\lambda(x) \\
 &= \int f(x)^t g(x)^{1-t} d\lambda(x) \\
 &= I(P, Q)(t)
 \end{aligned} \tag{4.13}$$

Hierbei wurde noch entscheidend (4.8) und (4.9) ausgenutzt.

Für die eben erzielten Ergebnisse gibt es noch naheliegende Verallgemeinerungen.

SATZ 4.3

$(Y, \mathcal{A})$ sei ein Maßraum

$$T : X \rightarrow Y$$

sei $(\mathcal{X} - \mathcal{M})$ - meßbar, und $\mathcal{P}$ sei eine durch λ dominierte Familie von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$. Für je zwei Maße $P, Q \in \mathcal{P}$ gebe es eine Umgebung von $t_0 = 1$, so daß dort gilt:

$$(i) \quad \frac{d}{dt} [I(P, Q)(t)] = \int f^t g^{1-t} \log \frac{f}{g} d\lambda \tag{4.14}$$

$$(ii) \quad \frac{d}{dt} [I(T(P), T(Q))(t)] = \int E_\lambda(f|y)^t E_\lambda(g|y)^{1-t} \log \frac{E_\lambda(f|y)}{E_\lambda(g|y)} dT(\lambda)(y) \tag{4.15}$$

$$(iii) \quad I(P, Q)(t) = I(T(P), T(Q))(t) \tag{4.16}$$

Dann gilt: T ist suffizient für $\mathcal{P}$.

Beweis:

Er folgt direkt aus Satz 4.1 und der Tatsache, daß für dominierte Klassen von Wahrscheinlichkeitsmaßen paarweise Suffizienz und Suffizienz äquivalent sind. (Siehe Halmos-Savage [19], S. 239, Theorem 3.)

Analog zu Satz 4.2 jetzt

Satz 4.4

$(Y, \mathcal{Y})$ sei ein Maßraum

$$T : X \rightarrow Y$$

sei $(\mathcal{X} - \mathcal{M})$ - meßbar, und $\mathcal{P}$ sei eine durch λ dominierte Familie von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$. Für je zwei Maße $P, Q \in \mathcal{P}$ gelte: T ist suffizient für $\{P, Q\}$. Dann gilt auch stets für alle $t \in [0, 1]$:

$$I(P, Q)(t) = I(T(P), T(Q))(t) \quad (4.17)$$

Beweis:

Folgt direkt aus Satz 4.2 und der gleichen Argumentation wie im Beweis zu Satz 4.3.

Zur Illustration nun einige Beispiele:

Beispiel 1:

Wir betrachten $X = \{(0,0), (0,1), (1,0), (1,1)\} \subset \mathbb{R}^2$

$$\mathcal{X} = \mathcal{P}(X), \quad p \in (0,1)$$

P, Q seien Wahrscheinlichkeitsmaße auf $\mathcal{X}$, die wie folgt festgelegt sind:

$$P(0,1) = P(1,0) = \frac{p}{2}$$

$$P(0,0) = P(1,1) = \frac{1-p}{2}$$

$$Q(0,1) = Q(0,0) = Q(1,1) = Q(1,0) = \frac{1}{4} \quad (4.18)$$

Dann ist die Statistik

$$T : X \rightarrow \mathbb{R}$$

$$(x,y) \rightarrow T(x,y) = \begin{cases} 1 & \text{falls } x = y \\ 0 & \text{falls } x \neq y \end{cases} \quad (4.19)$$

suffizient für $\{P, Q\}$.

Zunächst ist nämlich

$$\begin{aligned} I(P, Q)(t) &= \sum_{x=0}^1 \sum_{y=0}^1 P(x,y)^t Q(x,y)^{1-t} \\ &= 2 \left(\frac{1-p}{2} \right)^t \left(\frac{1}{4} \right)^{1-t} + 2 \left(\frac{p}{2} \right)^t \left(\frac{1}{4} \right)^{1-t} \end{aligned}$$

$$\begin{aligned}
 &= 2 \left(\frac{1}{4} \right)^{1-t} \left[\left(\frac{1-p}{2} \right)^t + \left(\frac{p}{2} \right)^t \right] \\
 &= \left(\frac{1}{2} \right)^{1-t} \left[(1-p)^t + p^t \right]
 \end{aligned} \tag{4.20}$$

und

$$I(T(P), T(Q))(t) = \sum_{\tau=0}^1 P(T=\tau)^t Q(T=\tau)^{1-t} \tag{4.21}$$

Nun gilt

$$\begin{aligned}
 P(T=0) &= p & ; & & Q(T=0) &= \frac{1}{2} \\
 P(T=1) &= 1-p & ; & & Q(T=1) &= \frac{1}{2}
 \end{aligned} \tag{4.22}$$

Damit ist

$$\begin{aligned}
 I(T(P), T(Q))(t) &= p^t \left(\frac{1}{2} \right)^{1-t} + (1-p)^t \left(\frac{1}{2} \right)^{1-t} \\
 &= \left(\frac{1}{2} \right)^{1-t} (p^t + (1-p)^t)
 \end{aligned} \tag{4.23}$$

Es sind alle Voraussetzungen von Satz 4.1 erfüllt. T ist mithin paarweise suffizient für $\{P, Q\}$.

Bemerkung:

Die Suffizienz von T erkennt man auch direkt mit Hilfe des Faktorisations-theorems für suffiziente Statistiken (siehe auch Witting [56] S. 149), wenn man

$$h_P : \{0,1\} \rightarrow \mathbb{R}$$

durch

$$\begin{aligned}
 h_P(0) &:= p \\
 h_P(1) &:= 1-p
 \end{aligned} \tag{4.24}$$

und

$$\begin{aligned}
 h_Q(0) &:= \frac{1}{2} \\
 h_Q(1) &:= \frac{1}{2}
 \end{aligned} \tag{4.25}$$

definiert.

Dann gilt:

$$P(x, y) = h_P(T(x, y)) \tag{4.26}$$

$$Q(x, y) = h_Q(T(x, y)) \tag{4.27}$$

für alle $(x, y) \in X$.

Beispiel 2:

Seien $X_1, \dots, X_k$ unabhängige Zufallsvariablen mit der Dichtefunktion

$$P_n(X_i = x) = \frac{1}{n} \quad (4.28)$$

$x = 1, \dots, n$

$i = 1, \dots, k$

d.h. $X = \{1, \dots, n, \dots\}$, $\mathcal{X} = \mathcal{P}(X)$

Der Träger der Dichtefunktion von P_n ist mithin die Menge $\{1, \dots, n\}$.

Dann ist

$$T := \max_{1 \leq i \leq k} X_i \quad (4.29)$$

eine suffiziente Statistik für $\{P_n\}_{n \in \mathbb{N}}$.

$\{P_n\}_{n \in \mathbb{N}}$ ist als abzählbare Menge von Wahrscheinlichkeitsmaßen dominiert

(z.B. durch $\sum_{n=1}^{\infty} \frac{1}{2^n} P_n$ oder das abzählende Maß auf $\mathbb{N}$).

Es genügt also der Nachweis der paarweisen Suffizienz. Seien also

P_n, P_m zwei beliebige Maße der Menge $\{P_n\}_{n \in \mathbb{N}}$, wobei ohne Einschränkung $m > n$ sei. Wie man schnell nachrechnet, gilt

$$I(P_n, P_m) = \left(\frac{n}{m}\right)^{1-t} \quad (4.30)$$

Bezeichnet $P_n^{(k)}$ die gemeinsame Verteilung der Zufallsvariablen $X_1, \dots, X_k$, so ist wegen der Unabhängigkeit (siehe Kapitel 5):

$$I(P_n^{(k)}, P_m^{(k)})(t) = \left(\frac{n}{m}\right)^{k(1-t)} \quad (4.31)$$

Außerdem gilt:

$$\begin{aligned} P_n(T = y) &= P_n(T \leq y) - P_n(T > y) \\ &= \frac{y^k}{n^k} - \frac{(y-1)^k}{n^k} \end{aligned} \quad (4.32)$$

$y = 1, \dots, n$

Dann folgt:

$$\begin{aligned} I(T(P_n), T(P_m))(t) &= \sum_{y=1}^n \left[\frac{y^k - (y-1)^k}{n^k} \right]^t \left[\frac{y^k - (y-1)^k}{m^k} \right]^{1-t} \\ &= n^{-kt} m^{-k(1-t)} \sum_{y=1}^n (y^k - (y-1)^k) \\ &= n^{-kt} m^{-k(1-t)} n^k \end{aligned}$$

$$= \left(\frac{n}{m} \right)^{k(1-t)} \quad (4.33)$$

Also hat man

$$I(P_n^{(k)}, P_m^{(k)})(t) = I(T(P_n), T(P_m))(t) \quad (4.34)$$

für alle $t \in [0, 1]$. Wegen Satz 4.3 ist dann $T = \max_{1 \leq i \leq k} X_i$ suffizient für $\{P_n\}_{n \in \mathbb{N}}$.

V. Kapitel

Produkträume

Wir untersuchen jetzt die Verträglichkeit von AEF und der Produktbildung von Wahrscheinlichkeitsräumen. Dazu wird das folgende Lemma benötigt. In der Terminologie wird sich Bauer [2] angeschlossen.

LEMMA 5.1

$(X_i, \mathfrak{X}_i, \mu_i, \nu_i)_{i=1}^n$ sei eine Familie von σ -endlichen Maßräumen. Dann gelten die folgenden Beziehungen:

$$(i) \quad \bigwedge_{i=1, \dots, n} \mu_i \ll \nu_i \Rightarrow \bigotimes_{i=1}^n \mu_i \ll \bigotimes_{i=1}^n \nu_i \quad (5.1)$$

und

$$\frac{d\left(\bigotimes_{i=1}^n \mu_i\right)}{d\left(\bigotimes_{i=1}^n \nu_i\right)} = \prod_{i=1}^n \frac{d\mu_i}{d\nu_i} \quad (5.2)$$

$$(ii) \quad \bigwedge_{i=1, \dots, n} \mu_i \equiv \nu_i \Rightarrow \bigotimes_{i=1}^n \mu_i \equiv \bigotimes_{i=1}^n \nu_i \quad (5.3)$$

Beweis:

Siehe Hewitt-Stromberg [23], S. 394.

Der folgende wichtige Satz findet Anwendung vornehmlich im letzten Kapitel.

SATZ 5.2

$(X_i, \mathfrak{X}_i, P_i, Q_i)_{i=1}^n$ sei eine Familie von Wahrscheinlichkeitsräumen. Weiterhin gelte $P_i, Q_i \ll \lambda_i$ ($i = 1, \dots, n$) mit σ -endlichem λ_i, λ_i Maß auf $\mathfrak{X}_i$. Dann gilt:

$$I\left(\bigotimes_{i=1}^n P_i, \bigotimes_{i=1}^n Q_i\right)(t) = \prod_{i=1}^n I(P_i, Q_i)(t) \quad (5.4)$$

Beweis:

Wir setzen per definitionem:

$$f^{\otimes} := \frac{d \left(\bigotimes_{i=1}^n P_i \right)}{d \left(\bigotimes_{i=1}^n \lambda_i \right)} \quad (5.5)$$

$$g^{\otimes} := \frac{d \left(\bigotimes_{i=1}^n Q_i \right)}{d \left(\bigotimes_{i=1}^n \lambda_i \right)} \quad (5.6)$$

Dann ist

$$\begin{aligned} I \left(\bigotimes_{i=1}^n P_i, \bigotimes_{i=1}^n Q_i \right)(t) &= \int [f^{\otimes}]^t [g^{\otimes}]^{1-t} d \left(\bigotimes_{i=1}^n \lambda_i \right) \\ &= \int \left[\prod_{i=1}^n f_i \right]^t \left[\prod_{i=1}^n g_i \right]^{1-t} d \left(\bigotimes_{i=1}^n \lambda_i \right) \\ &= \int \prod_{i=1}^n [f_i^t g_i^{1-t}] d \left(\bigotimes_{i=1}^n \lambda_i \right) \\ &= \prod_{i=1}^n \int f_i^t g_i^{1-t} d\lambda_i \\ &= \prod_{i=1}^n I(P_i, Q_i)(t) \end{aligned} \quad (5.7)$$

wobei

$$f_{i.} := \frac{dP_i}{d\lambda_i} \quad (5.8)$$

$$g_i := \frac{dQ_i}{d\lambda_i} \quad (5.9)$$

gesetzt und Lemma 5.1 ausgenützt wurde.

KOROLLAR 5.3

Es mögen dieselben Voraussetzungen wie in Satz 5.2 gelten, nur sei

$$X_i = X, \quad \mathcal{X}_i = \mathcal{X}, \quad P_i = P, \quad Q_i = Q, \quad \lambda_i = \lambda$$

($i = 1, \dots, n$)

Dann gilt für alle $t \in [0, 1]$

$$I(P^n, Q^n)(t) = [I(P, Q)(t)]^n \quad (5.10)$$

wenn man

$$P^n := \bigotimes_{i=1}^n P_i, \quad Q^n := \bigotimes_{i=1}^n Q_i \quad (5.11)$$

setzt.

KOROLLAR 5.4

Voraussetzungen wie in Satz 5.2, und sei $E_i \in \mathcal{X}$ für $i = 1, \dots, n$.

Dann zeigt sich analog zum dortigen Beweisschema:

$$I_{\bigotimes_{i=1}^n \mathcal{X}_{E_i}} \left(\bigotimes_{i=1}^n P_i, \bigotimes_{i=1}^n Q_i \right)(t) = \prod_{i=1}^n I_{E_i}(P_i, Q_i)(t) \quad (5.12)$$

Eine leichte Folgerung ist jetzt

SATZ 5.5

Sei $X_i = \mathbb{R}^p$, $\mathcal{X}_i = \mathcal{L}^p$ (Borelscher Mengenkörper auf $\mathbb{R}^p$), P_i, Q_i seien Wahrscheinlichkeitsmaße auf $\mathcal{X}_i$ ($i = 1, \dots, n$). Auf $(X, \mathcal{X})$ mit

$$X = \mathbb{R}^{pn} \quad (5.13)$$

$$\mathcal{X} = \mathcal{L}^{pn} \quad (5.14)$$

werden $\bigcirc_{i=1}^n P_i$, $\bigcirc_{i=1}^n Q_i$ betrachtet. Dann sind

$$\bigstar_{i=1}^n P_i := A_n \left(\bigcirc_{i=1}^n P_i \right) \quad (5.15)$$

$$\bigstar_{i=1}^n Q_i := A_n \left(\bigcirc_{i=1}^n Q_i \right) \quad (5.16)$$

die entsprechenden Faltungen auf $(\mathbb{R}^P, \mathcal{G}^P)$.
(Vgl. Bauer [2], S. 104.)

Dabei ist

$$A_n : (\mathbb{R}^P)^n \rightarrow \mathbb{R}^P$$

$$(x_1, \dots, x_n) \rightarrow A_n(x_1, \dots, x_n) := \sum_{i=1}^n x_i \quad (5.17)$$

Dann gilt:

$$\prod_{i=1}^n I(P_i, Q_i)(t) \leq I \left(\bigstar_{i=1}^n P_i, \bigstar_{i=1}^n Q_i \right)(t) \quad (5.18)$$

Beweis:

Er folgt unmittelbar aus Satz 1.8 und Satz 5.2.

VI. Kapitel

Grenzwertsätze

Die nachstehenden Sätze haben ihre Analoga für das Abstandsmaß $K(P, Q)$ von Kullback-Leibler (siehe Kullback [40]).

SATZ 6.1

$(Y, \mathcal{M})$ sei ein Maßraum mit separablem $\mathcal{M}$. Für jedes $n \in \mathbb{N}$ sei

$$T_n : X \rightarrow Y$$

$(\mathcal{X} - \mathcal{M})$ - meßbare Transformation. Weiterhin sei für $F \in \mathcal{M}$:

$$\lim_{n \rightarrow \infty} T_n(P)(F) = T(P)(F) \quad (6.1)$$

$$\lim_{n \rightarrow \infty} T_n(Q)(F) = T(Q)(F) \quad (6.2)$$

wobei

$$T : X \rightarrow Y$$

ebenfalls eine $(\mathcal{X} - \mathcal{M})$ - meßbare Transformation sei. Dann gilt für alle $t \in [0, 1]$:

$$\limsup_{n \rightarrow \infty} I(T_n(P), T_n(Q))(t) \leq (T(P), T(Q))(t) \quad (6.3)$$

Beweis:

Wie gezeigt wurde, gilt für $n \in \mathbb{N}$ und $t \in [0, 1]$:

$$I(T_n(P), T_n(Q))(t) \leq \sum_{i=1}^k T_n(P)(F_i)^t T_n(Q)(F_i)^{1-t} \quad (6.4)$$

wobei $\{F_i\}_{i=1}^k$ eine beliebige endliche Partition von Y ist. Bildet man auf beiden Seiten von (6.4) den Limes superior, so ist für alle $t \in [0, 1]$:

$$\limsup_{n \rightarrow \infty} I(T_n(P), T_n(Q))(t) \leq \sum_{i=1}^k T(P)(F_i)^t T(Q)(F_i)^{1-t} \quad (6.5)$$

Da $\mathcal{M}$ separabel ist, gilt wegen Satz 2.9

$$I(T(P), T(Q))(t) = \inf_{\mathcal{F} = \{F_1, \dots, F_m\} \text{ endliche Partition von } Y} \sum_{i=1}^m T(P)(F_i)^t T(Q)(F_i)^{1-t} \quad (6.6)$$

Damit folgt schon die Behauptung.

SATZ 6.2

$\mathfrak{X}$ sei separabel und $\{P_n\}_{n \in \mathbb{N}}$ sei eine Folge von Wahrscheinlichkeitsmaßen auf $\mathfrak{X}$ mit

$$\lim_{n \rightarrow \infty} P_n(E) = P(E) \quad \text{für alle } E \in \mathfrak{X} \quad (6.7)$$

Dann gilt für alle $t \in [0, 1]$:

$$\limsup_{n \rightarrow \infty} I(P_n, Q)(t) \leq I(P, Q)(t) \quad (6.8)$$

Beweis:

Analog zum Beweis von Satz 6.1.

SATZ 6.3

$\mathfrak{X}$ sei separabel, $\{P_n, Q_n\}_{n \in \mathbb{N}}$ sei eine Doppelfolge von Wahrscheinlichkeitsmaßen auf $\mathfrak{X}$ mit

$$\lim_{n \rightarrow \infty} P_n(E) = P(E) \quad (6.9)$$

$$\lim_{n \rightarrow \infty} Q_n(E) = Q(E) \quad (6.10)$$

für alle $E \in \mathfrak{X}$. Dann gilt für alle $t \in [0, 1]$:

$$\limsup_{n \rightarrow \infty} I(P_n, Q_n)(t) \leq I(P, Q)(t) \quad (6.11)$$

Beweis:

Analog zum Beweis von Satz 6.1.

Für Satz 6.3 wählen wir folgendes Beispiel, das auch bei Kullback [40] in ähnlichem Zusammenhang auftaucht.

Beispiel:

Für P wähle man die Poisson-Verteilung mit dem Parameter λ_1 .

Für Q wähle man ebenfalls die Poisson-Verteilung mit dem Parameter λ_2 .

$(p_n)_{n=1}^{\infty}$ und $(q_n)_{n=1}^{\infty}$ seien zwei reelle Folgen mit

$$0 < p_n < 1, \quad 0 < q_n < 1 \quad (6.12)$$

und

$$\lim_{n \rightarrow \infty} np_n = \lambda_1 \quad (6.13)$$

$$\lim_{n \rightarrow \infty} nq_n = \lambda_2 \quad (6.14)$$

P_n sei die Binomialverteilung mit den Parametern (n, p_n) , Q_n sei die Binomialverteilung mit den Parametern (n, q_n) . Bekanntlich ist in diesen Fällen

$$X = \mathbb{N} \cup \{0\}, \quad \mathcal{X} = \mathcal{P}(X)$$

$$\lim_{n \rightarrow \infty} P_n(E) = P(E) \quad (6.15)$$

$$\lim_{n \rightarrow \infty} Q_n(E) = Q(E) \quad (6.16)$$

für alle $E \in \mathcal{X}$.

Es ist (siehe Anhang):

$$I(P_n, Q_n)(t) = [p_n^t q_n^{1-t} + (1-p_n)^t (1-q_n)^{1-t}]^n \quad (6.17)$$

$$I(P, Q)(t) = e^{-\lambda_1 t - \lambda_2(1-t) + \lambda_1^t \lambda_2^{1-t}} \quad (6.18)$$

Setzt man

$$\lambda_{n1} : = np_n \quad (6.19)$$

$$\lambda_{n2} : = nq_n \quad (6.20)$$

so wird

$$\begin{aligned} I(P_n, Q_n)(t) &= \left[\left(\frac{\lambda_{n1}}{n} \right)^t \left(\frac{\lambda_{n2}}{n} \right)^{1-t} + \left(1 - \frac{\lambda_{n1}}{n} \right)^t \left(1 - \frac{\lambda_{n2}}{n} \right)^{1-t} \right]^n \\ &= \left[\frac{\lambda_{n1}^t \lambda_{n2}^{1-t} + (n - \lambda_{n1})^t (n - \lambda_{n2})^{1-t}}{n} \right]^n \end{aligned} \quad (6.21)$$

Wir setzen nun

$$x_n : = \lambda_{n1}^t \lambda_{n2}^{1-t} + (n - \lambda_{n1})^t (n - \lambda_{n2})^{1-t} - n \quad (6.22)$$

Dann ist

$$I(P_n, Q_n)(t) = \left(1 + \frac{x_n}{n} \right)^n \quad (6.23)$$

Nun gilt wegen Erwe [12] S. 159

$$\begin{aligned} x_n &\leq \lambda_{n1}^t \lambda_{n2}^{1-t} + t(n - \lambda_{n1}) + (1-t)(n - \lambda_{n2}) - n \\ &\leq \lambda_{n1}^t \lambda_{n2}^{1-t} - \lambda_{n1}^t - \lambda_{n2}^{1-t} \end{aligned} \quad (6.24)$$

Außerdem ist

$$\left(1 + \frac{x_n}{n}\right) \geq 0 \quad (6.25)$$

und somit

$$\left(1 + \frac{x_n}{n}\right)^n \leq \left(1 + \frac{y_n}{n}\right)^n \quad (6.26)$$

mit

$$y_n := \lambda_{n1}^t \lambda_{n2}^{1-t} - \lambda_{n1}^t - \lambda_{n2}^{1-t} \quad (6.27)$$

Nun hat man wegen Fichtenholz [14] Bd. I S. 150

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 + \frac{y_n}{n}\right)^n &= e^{\lim_{n \rightarrow \infty} y_n} \\ &= I(P, Q)(t) \end{aligned} \quad (6.28)$$

Deshalb folgt

$$\limsup_{n \rightarrow \infty} \left(1 + \frac{x_n}{n}\right)^n \leq I(P, Q)(t) \quad (6.29)$$

oder

$$\limsup_{n \rightarrow \infty} I(P_n, Q_n)(t) \leq I(P, Q)(t) \quad (6.30)$$

Wie man Satz 1.5 entnimmt, sind für eine Folge von Wahrscheinlichkeitsmaßen $\{P_n\}_{n \in \mathbb{N}}$ für alle $t \in (0, 1)$ die beiden Aussagen

$$\lim_{n \rightarrow \infty} I(P_n, Q)(t) = 1 \quad (6.31)$$

$$\lim_{n \rightarrow \infty} \int |f_n - g| d\lambda = 0 \quad (6.32)$$

äquivalent, wobei

$$f_n := \frac{dP_n}{d\lambda} \quad (6.33)$$

für alle $n \in \mathbb{N}$.

Im allgemeinen folgt aus (6.32) bekanntlich nicht, daß

$$\lim_{n \rightarrow \infty} f_n = g \quad [\lambda] \quad (6.34)$$

Für diskrete Wahrscheinlichkeitsräume gilt immerhin

SATZ 6.4

$X = \{x_1, \dots, x_m\}$ sei ein diskreter endlicher Meßraum (versehen mit der σ -Algebra $\mathcal{X} = \mathcal{O}(X)$). Sei

$$P(x_i) = p_i > 0 \quad (6.35)$$

$$Q_n(x_i) = q_{in} > 0 \quad (6.36)$$

wobei $i = 1, \dots, m, n \in \mathbb{N}$.

Es sei für ein $t_0 \in (0, 1)$:

$$\lim_{n \rightarrow \infty} I(P, Q_n)(t_0) = 1 \quad (6.37)$$

Dann gilt für $i = 1, \dots, m$

$$\lim_{n \rightarrow \infty} q_{in} = p_i \quad (6.38)$$

Beweis:

Folgt direkt aus Rényi [48] S. 483 und Satz 1.5. Es folgt das wichtigste Ergebnis des Kapitels:

SATZ 6.5

$\{P_n\}_{n \in \mathbb{N}}, \{Q_n\}_{n \in \mathbb{N}}$ seien zwei Folgen von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$ mit $P_n, Q_n \ll \lambda$ für alle $n \in \mathbb{N}$. Es sei

$$f_n := \frac{dP_n}{d\lambda} \quad ; \quad g_n := \frac{dQ_n}{d\lambda}$$

$$E_n := \text{TR}(f_n) \quad ; \quad E := \text{TR}(f)$$

Weiterhin wird vorausgesetzt für $t_0 \in [0, 1]$:

$$(i) \quad \text{TR}(f) = \text{TR}(g) \quad [\lambda] \quad (6.40)$$

$$(ii) \quad \text{TR}(f_n) = \text{TR}(g_n) \quad [\lambda] \quad \text{für alle } n \in \mathbb{N} \quad (6.41)$$

$$(iii) \quad TR(f_n) \subset TR(f) \quad [\lambda] \quad \text{für alle } n \in \mathbb{N} \quad (6.42)$$

$$(iv) \quad \lim_{n \rightarrow \infty} Q(E_n \Delta E) = 0 \quad (6.43)$$

$$(v) \quad \lim_{n \rightarrow \infty} \operatorname{ess\,sup}_{E_n} \left| 1 - \frac{g_n}{g} \right| = 0 \quad (6.44)$$

$$(vi) \quad \lim_{n \rightarrow \infty} \operatorname{ess\,sup}_{E_n} \left| \left(\frac{f}{g} \right)^{t_0} - \left(\frac{f_n}{g_n} \right)^{t_0} \right| = 0 \quad (6.45)$$

Dann gilt:

$$\lim_{n \rightarrow \infty} I(P_n, Q_n)(t_0) = I(P, Q)(t_0) \quad (6.46)$$

Beweis:

Es ist

$$|I(P, Q)(t_0) - I(P_n, Q_n)(t_0)| = \left| \int_E f^{t_0} g^{1-t_0} d\lambda - \int_{E_n} f_n^{t_0} g_n^{1-t_0} d\lambda \right| \quad (6.47)$$

$$= \left| \int_{E_n} f^{t_0} g^{1-t_0} d\lambda + \int_{E \setminus E_n} f^{t_0} g^{1-t_0} d\lambda - \int_{E_n} f_n^{t_0} g_n^{1-t_0} d\lambda \right| \quad (6.48)$$

$$\leq \left| \int_{E \setminus E_n} f^{t_0} g^{1-t_0} d\lambda \right| + \left| \int_{E_n} f^{t_0} g^{1-t_0} d\lambda - \int_{E_n} f_n^{t_0} g_n^{1-t_0} d\lambda \right| \quad (6.49)$$

Damit gilt

$$|I(P, Q)(t_0) - I(P_n, Q_n)(t_0)| \leq \int_{E \setminus E_n} \left(\frac{f}{g} \right)^{t_0} dQ + \left| \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ - \int_{E_n} \left(\frac{f_n}{g_n} \right)^{t_0} dQ_n \right| \quad (6.50)$$

Da nach Voraussetzung

$$\lim_{n \rightarrow \infty} Q(E \Delta E_n) = 0 \quad (6.51)$$

gilt auch

$$\lim_{n \rightarrow \infty} \int_{E \setminus E_n} \left(\frac{f}{g} \right)^{t_0} dQ = 0 \quad (6.52)$$

Weiterhin ist

$$\left| \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ - \int_{E_n} \left(\frac{f}{g_n} \right)^{t_0} dQ_n \right| =$$

$$\left| \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ + \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ_n - \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ_n - \int_{E_n} \left(\frac{f}{g_n} \right)^{t_0} dQ_n \right| \quad (6.53)$$

$$\leq \left| \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ - \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ_n \right| + \left| \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ_n - \int_{E_n} \left(\frac{f}{g_n} \right)^{t_0} dQ_n \right| \quad (6.54)$$

Wir setzen nun

$$A := \left| \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ - \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ_n \right| \quad (6.55)$$

$$B := \left| \int_{E_n} \left(\frac{f}{g} \right)^{t_0} dQ_n - \int_{E_n} \left(\frac{f}{g_n} \right)^{t_0} dQ_n \right| \quad (6.56)$$

und schätzen zunächst A nach oben ab.

$$A = \left| \int_{E_n} f^{t_0} g^{1-t_0} d\lambda - \int_{E_n} f^{t_0} g^{-t_0} g_n d\lambda \right|$$

$$= \left| \int_{E_n} f^{t_0} g^{-t_0} (g - g_n) d\lambda \right| \quad (6.57)$$

Weil $g = \frac{dQ}{d\lambda}$, ist

$$A = \left| \int_{E_n} f^{t_0} g^{-t_0} \left(\frac{g - g_n}{g} \right) dQ \right|$$

$$\leq \operatorname{ess\,sup}_{E_n} \left| 1 - \frac{g_n}{g} \right| \cdot \int_{E_n} f^{t_0} g^{-t_0} dQ$$

$$= \operatorname{ess\,sup}_{E_n} \left| 1 - \frac{g_n}{g} \right| \cdot I_{E_n}(P, Q)(t_0) \quad (6.58)$$

Für B hat man

$$\begin{aligned}
 B &= \left| \int_{E_n} \left[\left(\frac{f}{g} \right)^{t_0} - \left(\frac{f_n}{g_n} \right)^{t_0} \right] dQ_n \right| \\
 &\leq \operatorname{ess\,sup}_{E_n} \left| \left(\frac{f}{g} \right)^{t_0} - \left(\frac{f_n}{g_n} \right)^{t_0} \right| \cdot \int_{E_n} dQ_n \\
 &= \operatorname{ess\,sup}_{E_n} \left| \left(\frac{f}{g} \right)^{t_0} - \left(\frac{f_n}{g_n} \right)^{t_0} \right| Q_n(E_n)
 \end{aligned} \tag{6.59}$$

Da $I_{E_n}(P, Q)(t_0)$ und $Q_n(E_n)$ beschränkt sind, folgt aus (v) und (vi) schließlich die Behauptung.

Bemerkung:

Ersetzt man (vi) durch (vi)', so gilt die Behauptung immer noch, wobei

$$(vi)' \quad \lim_{n \rightarrow \infty} \operatorname{ess\,sup}_{E_n} \left| 1 - \left(\frac{f_n g}{g_n f} \right)^{t_0} \right| = 0 \tag{6.60}$$

Beweis:

Es genügt, B aus vorigem Theorem nach oben abzuschätzen.

$$\begin{aligned}
 B &= \left| \int_{E_n} \left[\left(\frac{f}{g} \right)^{t_0} - \left(\frac{f_n}{g_n} \right)^{t_0} \right] dQ_n \right| \\
 &= \left| \int_{E_n} \left[1 - \left(\frac{f_n g}{g_n f} \right)^{t_0} \right] \left[\frac{f}{g} \right]^{t_0} dQ_n \right| \\
 &\leq \int_{E_n} \left| 1 - \left(\frac{f_n g}{g_n f} \right)^{t_0} \right| \left(\frac{f}{g} \right)^{t_0} dQ_n \\
 &\leq \operatorname{ess\,sup}_{E_n} \left| 1 - \left(\frac{f_n g}{g_n f} \right)^{t_0} \right| \int_{E_n} f^{t_0} g^{-t_0} dQ_n \\
 &= \operatorname{ess\,sup}_{E_n} \left| 1 - \left(\frac{f_n g}{g_n f} \right)^{t_0} \right| \cdot \int_{E_n} f^{t_0} g^{-t_0} g_n \, d\lambda
 \end{aligned} \tag{6.61}$$

Wegen (v) gilt

$$\lim_{n \rightarrow \infty} \operatorname{ess\,sup}_{E_n} \left| 1 - \frac{g_n}{g} \right| = 0 \quad (6.62)$$

Sei jetzt $\varepsilon > 0$ beliebig. Für $n \geq N(\varepsilon)$ gibt es somit $E'_n \subset E_n$, $E'_n \in \mathcal{X}$ mit $\lambda(E'_n) = 0$ und

$$\left| 1 - \frac{g_n}{g}(x) \right| < \varepsilon \quad (6.63)$$

für alle $x \in E_n \setminus E'_n$.

Daraus folgt, daß

$$g_n < (1 + \varepsilon)g \quad [\lambda] \quad (6.64)$$

auf E_n .

Damit hat man für alle $n \geq N(\varepsilon)$

$$\int_{E_n} f \circ g^{-t_0} g_n \, d\lambda \leq (1 + \varepsilon) \int_{E_n} f \circ g^{1-t_0} \, d\lambda \quad (6.65)$$

Es ist also insgesamt

$$B \leq \operatorname{ess\,sup}_{E_n} \left| 1 - \left(\frac{f \circ g}{g_n f} \right)^{t_0} \right| (1 + \varepsilon) I_{E_n}(P, Q)(t_0) \quad (6.66)$$

für alle $n \geq N(\varepsilon)$.

Da $I_{E_n}(P, Q)(t_0)$ beschränkt ist, folgt aus (vi)' die Behauptung.

KOROLLAR 6.6

$\{P_n\}_{n \in \mathbb{N}}$ sei eine Folge von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$ mit $P_n \ll \lambda$ für alle $n \in \mathbb{N}$

$$f_n := \frac{dP_n}{d\lambda}$$

$$E := \operatorname{TR}(f) \quad (6.67)$$

Weiterhin gelte

$$(i) \quad \operatorname{TR}(f) = \operatorname{TR}(f_n) \quad [\lambda] \quad \text{für alle } n \in \mathbb{N} \quad (6.68)$$

(ii) Für $t_0 \in [0,1]$ sei

$$\lim_{n \rightarrow \infty} \operatorname{ess\,sup}_E \left| 1 - \left(\frac{f_n}{f} \right)^{t_0} \right| = 0 \quad (6.69)$$

Dann gilt

$$\lim_{n \rightarrow \infty} I(P_n, P)(t_0) = 1 \quad (6.70)$$

Beweis:

Folgt direkt aus Satz 6.5, wenn man $g = g_n = f$ wählt.

KOROLLAR 6.7

Voraussetzungen wie in Korollar 6.6. Dann gilt auch:

$$\lim_{n \rightarrow \infty} I(P, P_n)(t_0) = 1 \quad (6.71)$$

Beweis:

Ersetze im vorigen Korollar 6.7 t_0 durch $1 - t_0$. Man beachte, daß

$$I(P, Q)(t) = I(Q, P)(1 - t) \quad (6.72)$$

KOROLLAR 6.8

$\{P_n\}_{n \in \mathbb{N}}$ sei eine Folge von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$ mit $P_n \ll \lambda$ für alle $n \in \mathbb{N}$.

$$E := \operatorname{TR}(f)$$

$$f_n := \frac{dP_n}{d\lambda} \quad (6.73)$$

Weiterhin gelte für $t_0 \in [0,1]$:

$$(i) \quad \operatorname{TR}(f) = \operatorname{TR}(g) = \operatorname{TR}(f_n) \quad [\lambda] \quad (6.74)$$

für alle $n \in \mathbb{N}$

$$(ii) \quad \lim_{n \rightarrow \infty} \operatorname{ess\,sup} \left| 1 - \left(\frac{f_n}{f} \right)^{t_0} \right| = 0 \quad (6.75)$$

Dann gilt:

$$\lim_{n \rightarrow \infty} I(P_n, Q)(t_0) = I(P, Q)(t_0) \quad (6.76)$$

Beweis:

Folgt unmittelbar aus der Ergänzung (vi)' zu Satz 6.5, wenn man g_n durch g ersetzt.

KOROLLAR 6.9

$\{Q_n\}_{n \in \mathbb{N}}$ sei eine Folge von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$ mit $Q_n \ll \lambda$ für alle $n \in \mathbb{N}$.

$$E := TR(f)$$

$$g_n := \frac{dQ_n}{d\lambda} \quad (6.77)$$

Weiterhin gelte für $t_0 \in [0, 1]$:

$$(i) \quad TR(f) = TR(g) = TR(g_n) \quad [\lambda] \quad (6.78)$$

für alle $n \in \mathbb{N}$.

$$(ii) \quad \lim_{n \rightarrow \infty} \operatorname{ess\,sup}_E |1 - (\frac{g}{g_n})^{t_0}| = 0 \quad (6.79)$$

Dann gilt:

$$\lim_{n \rightarrow \infty} I(P, Q_n)(t_0) = I(P, Q)(t_0) \quad (6.80)$$

Beweis:

Folgt direkt aus der Ergänzung (vi)' zu Satz 6.5, wenn man f_n durch f ersetzt.

KOROLLAR 6.10

Sei $\{E_n\}_{n \in \mathbb{N}}$ eine Folge von Elementen aus $\mathcal{X}$ mit $E_n \subset E$ für alle $n \in \mathbb{N}$ und

$$\lim_{n \rightarrow \infty} Q(E \Delta E_n) = 0 \quad (6.81)$$

$$\lim_{n \rightarrow \infty} \operatorname{ess\,sup}_{E_n} \left(\frac{f}{g} \right) < \infty \quad (6.82)$$

Dann gilt für $t_0 \in [0, 1]$:

$$\lim_{n \rightarrow \infty} I(P^{\phi_n}, Q^{\phi_n})(t_0) = I(P, Q)(t_0) \quad (6.83)$$

mit

$$\phi_n := \chi_{E_n}, \quad E := TR(f) = TR(g) \quad (6.84)$$

Beweis:

Man setze

$$f_n : = \frac{f \cdot \chi_{E_n}}{P(E_n)} \quad (6.85)$$

$$g_n : = \frac{g \cdot \chi_{E_n}}{Q(E_n)} \quad (6.86)$$

Es ist nun

$$\left| 1 - \frac{g_n}{g} \right| = \left| 1 - \frac{\chi_{E_n}}{Q(E_n)} \right| \quad (6.87)$$

Da nach Voraussetzung

$$\lim_{n \rightarrow \infty} Q(E_n) = 1 \quad (6.88)$$

so folgt:

$$\lim_{n \rightarrow \infty} \operatorname{ess\,sup}_{E_n} \left| 1 - \frac{g_n}{g} \right| = 0 \quad (6.89)$$

Außerdem wird

$$\left| \left(\frac{f}{g} \right)^{t_0} - \left(\frac{f_n}{g_n} \right)^{t_0} \right| = \left(\frac{f}{g} \right)^{t_0} \left| 1 - \frac{Q(E_n)}{P(E_n)} \right| \quad (6.90)$$

und weil

$$\lim_{n \rightarrow \infty} P(E_n) = \lim_{n \rightarrow \infty} Q(E_n) = 1 \quad (6.91)$$

hat man

$$\lim_{n \rightarrow \infty} \operatorname{ess\,sup}_{E_n} \left| \left(\frac{f}{g} \right)^{t_0} - \left(\frac{f_n}{g_n} \right)^{t_0} \right| = 0 \quad (6.92)$$

Wegen Satz 6.5 folgt die Behauptung.

VII. Kapitel

Anwendungen auf die Testtheorie

Dem Kapitel wird ein zentrales Lemma vorangestellt, das bekanntlich von Neyman-Pearson stammt. Den Beweis entnimmt man z.B. dem Buch von Hogg-Craig [25] S. 262.

LEMMA 7.1

Sei $E \in \mathcal{X}$ und $k \in \mathbb{R}^+$, so daß

$$\text{für } x \in E : k f(x) \leq g(x) \quad (7.1)$$

$$\text{für } x \in X \setminus E : k f(x) > g(x) \quad (7.2)$$

Dann gilt für alle $E' \in \mathcal{X}$ mit

$$P(E) = P(E') : \quad (7.3)$$

$$Q(E') \leq Q(E) \quad (7.4)$$

Dieses E heißt kritisches Gebiet.

Fundamental für das Weitere ist das nachstehende Lemma, das sich auf Lemma 7.1 stützt.

LEMMA 7.2

Es gebe ein $E \in \mathcal{X}$, welches die Voraussetzungen von Lemma 7.1 erfüllt.

Dann gilt für alle $t \in [0, 1]$:

$$(i) \quad P(E) \leq \left(\frac{1}{k} \right)^{1-t} I(P, Q)(t) \quad (7.5)$$

$$(ii) \quad 1 - Q(E) \leq k^t I(P, Q)(t) \quad (7.6)$$

wobei k wie in Lemma 7.1 gewählt sei.

Beweis:

Sei $t \in [0, 1]$ beliebig. Zunächst gilt, da

$$f(x) = f(x)^t f(x)^{1-t} \quad (7.7)$$

für $x \in E$:

$$f(x) \leq f(x)^t \left(\frac{g(x)}{k} \right)^{1-t} \quad (7.8)$$

Hieraus folgt:

$$\int_E f(x) d\lambda(x) \leq \int_E f(x) \left(\frac{g(x)}{k} \right)^{1-t} d\lambda(x) \quad (7.9)$$

Mithin gilt

$$P(E) \leq \left(\frac{1}{k}\right)^{1-t} I_E(P,Q)(t) \quad (7.10)$$

Da aber

$$I_E(P,Q)(t) \leq I(P,Q)(t) \quad (7.11)$$

folgt (i).

Ist $x \in X \setminus E$, so folgt

$$g^t(x) \leq k^t f^t(x) \quad (7.12)$$

und damit

$$g(x) \leq k^t f^t(x) g^{1-t}(x) \quad (7.13)$$

Integriert man über $\bar{E} = X \setminus E$, so erhält man:

$$Q(\bar{E}) \leq k^t I_{\bar{E}}(P,Q)(t) \quad (7.14)$$

oder anders geschrieben

$$1 - Q(E) \leq k^t I_E(P,Q)(t) \quad (7.15)$$

Da aber, analog zu oben,

$$I_{\bar{E}}(P,Q)(t) \leq I(P,Q)(t) \quad (7.16)$$

folgt das Desideratum

$$1 - Q(E) \leq k^t I(P,Q)(t) \quad (7.17)$$

Man braucht jetzt die

DEFINITION

$\mathcal{P}, \mathcal{Q}$ seien zwei Familien von Wahrscheinlichkeitsmaßen auf $\mathcal{X}$,
und sei $t \in [0,1]$. Dann heißt

$$d_t(\mathcal{P}, \mathcal{Q}) := \sup_{(P,Q) \in \mathcal{P} \times \mathcal{Q}} I(P,Q)(t) \quad (7.18)$$

der t -Abstand von $\mathcal{P}$ und $\mathcal{Q}$.

Bemerkung:

Es ist stets

$$0 \leq d_t(\mathcal{P}, \mathcal{Q}) \leq 1 \quad (7.19)$$

Mit Hilfe des Begriffes t -Abstand zweier Familien von Wahrscheinlichkeits-
maßen läßt sich dann das folgende Lemma formulieren:

LEMMA 7.3

Wir betrachten die beiden Hypothesen

$$H : (X_1, \dots, X_n) \text{ ist verteilt nach } P \in \mathcal{P} \quad (7.20)$$

$$G : (X_1, \dots, X_n) \text{ ist verteilt nach } Q \in \mathcal{Q} \quad (7.21)$$

wobei $\mathcal{P}, \mathcal{Q}$ Familien von Wahrscheinlichkeitsmaßen auf dem Stichprobenraum $(X, \mathcal{X})$ seien. $((X_1, \dots, X_n)$ ist als Stichprobe vom Umfang n aufzufassen).

Sei $t \in [0, 1]$ mit $d_t(\mathcal{P}, \mathcal{Q}) > 0$. Außerdem gelte:

(i) Es gibt $P_0 \in \mathcal{P}$, $Q_0 \in \mathcal{Q}$ mit ¹⁾

$$d_t(\mathcal{P}, \mathcal{Q}) = I(P_0, Q_0)(t) \quad (7.22)$$

(ii) Ist f_0 bzw. g_0 die Dichtefunktion von P_0 bzw. Q_0 , so gibt es eine Menge $W \in \mathcal{X}$ und ein $k \in \mathbb{R}^+$, so daß

$$W = \{(x_1, \dots, x_n) | k f_0(x_1, \dots, x_n) \leq g_0(x_1, \dots, x_n)\} \quad (7.23)$$

und

$$P(W) \leq P_0(W) \quad \text{für alle } P \in \mathcal{P} \quad (7.24)$$

$$Q(W) \geq Q_0(W) \quad \text{für alle } Q \in \mathcal{Q} \quad (7.25)$$

Dann gilt:

a) W ist kritisches Gebiet zum zusammengesetzten Test H gegen G .

$$b) \quad P(W) \leq \left(\frac{1}{k}\right)^{1-t} I(P_0, Q_0)(t) \quad \text{für alle } P \in \mathcal{P} \quad (7.26)$$

$$c) \quad 1 - Q(W) \leq k^t I(P_0, Q_0)(t) \quad \text{für alle } Q \in \mathcal{Q} \quad (7.27)$$

Beweis:

a) folgt direkt aus (ii). (Man vergleiche hierzu auch die statistische Literatur.)

b) und c) folgen unmittelbar aus (i) und Lemma 7.2.

Wenn die Voraussetzungen des vorigen Lemmas erfüllt sind, wird also die Wahrscheinlichkeit für den Fehler 1. Art unterhalb von

$$\left(\frac{1}{k}\right)^{1-t} I(P_0, Q_0)(t) \quad (7.28)$$

1) P_0 und Q_0 können allgemein von t abhängen. Dieser Fall wird im folgenden jedoch ausgeschlossen.

liegen, die Wahrscheinlichkeit, den Fehler 2. Art zu begehen, unterhalb von

$$k^t I(P_0, Q_0)(t) \quad (7.29)$$

In der Praxis ist man bei vorgegebener Irrtumswahrscheinlichkeit α bestrebt, das k so zu bestimmen, daß

$$P(W) \leq \alpha \quad \text{für } P \in \mathcal{P} \quad (7.30)$$

$$1 - Q(W) = \inf_W \quad \text{für } Q \in \mathcal{Q} \quad (7.31)$$

Dieser Test wird jetzt modifiziert, indem man k aus den folgenden Relationen ermittelt:

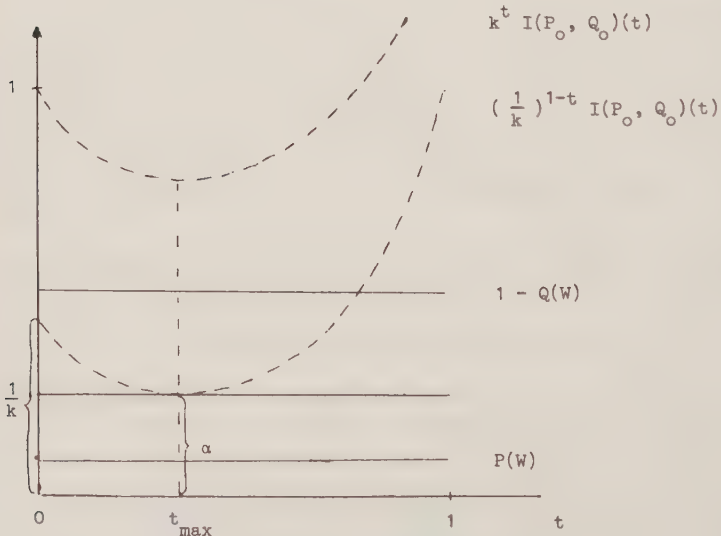
$$(i) \quad k f_0(x_1, \dots, x_n) \leq g_0(x_1, \dots, x_n) \quad (7.32)$$

(ii) Bestimme t so, daß

$$k^t I(P_0, Q_0)(t) = \min_t \quad (7.33)$$

unter der Nebenbedingung

$$\left(\frac{1}{k}\right)^{1-t} I(P_0, Q_0)(t) = \alpha \quad (7.34)$$



Zur Illustration dieses neuen Testprinzips dienen die folgenden drei Beispiele, die allesamt Testprobleme bei normalverteilter Grundgesamtheit behandeln.

Beispiel 1:

$(X_1, \dots, X_n)$ sei die Stichprobe einer normalverteilten Grundgesamtheit $(X_i$ unabhängig) mit unbekanntem μ und bekanntem σ . Das Testproblem laute:

$$H: \mu \leq -a \quad (7.35)$$

$$G: \mu \geq a \quad (7.36)$$

wobei $a > 0$ eine feste, vorgegebene Konstante ist. Trifft H zu, so ist

$$f(x_1, \dots, x_n | \mu_1, \sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2 \right\} \quad (7.37)$$

mit $\mu_1 \leq -a$.

Trifft jedoch G zu, so wird

$$g(x_1, \dots, x_n | \mu_2, \sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_2)^2 \right\} \quad (7.38)$$

mit $\mu_2 \geq a$.

$\mathcal{P}$ ist hier die Menge aller unter H zulässigen Wahrscheinlichkeitsverteilungen (genauer: bei Normalverteilung) über dem Stichprobenraum X , d.h.

$$\mathcal{P} = \left\{ P \mid \frac{dP}{d\lambda^n}(\mathcal{U}) = f(\mathcal{U} | \mu_1, \sigma), \mu_1 \leq -a \right\} \quad (7.39)$$

und

$$\mathcal{Q} = \left\{ Q \mid \frac{dQ}{d\lambda^n}(\mathcal{U}) = g(\mathcal{U} | \mu_2, \sigma), \mu_2 \geq a \right\} \quad (7.40)$$

die Menge aller unter G zulässigen Wahrscheinlichkeitsverteilungen über X (λ^n ist das Lebesgue-Maß über $\mathcal{X}^n$).

Nun gilt (vgl. Anhang) für $P \in \mathcal{P}$, $Q \in \mathcal{Q}$:

$$I(P, Q)(t) = \exp \left\{ -\frac{n}{2\sigma^2} t(1-t)(\mu_1 - \mu_2)^2 \right\} \quad (7.41)$$

Bei festem $t \in [0, 1]$ wird $I(P, Q)(t)$ maximal, wenn $(\mu_1 - \mu_2)^2$ minimal wird. Wählt man also

$$\mu_1 = -a \quad ; \quad \mu_2 = a \quad (7.42)$$

und

$$\frac{dP_o}{d\lambda^n} = f(\mathcal{U} | -a, \sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i + a)^2 \right\} \quad (7.43)$$

$$\frac{dQ_o}{d\lambda^n} = g(\mathcal{U} | a, \sigma) = \left(\frac{1}{\sqrt{\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2 \right\} \quad (7.44)$$

so gilt

$$d_t(\mathcal{P}, \mathcal{Q}) = I(P_o, Q_o)(t) \quad (7.45)$$

mit

$$I(P_o, Q_o)(t) = \exp \left\{ -\frac{2n}{\sigma^2} t(1-t) a^2 \right\} > 0 \quad (7.46)$$

Wir kommen nun zur Bestimmung der Konstanten k :

$$k f(\mathcal{U} | -a, \sigma) \leq g(\mathcal{U} | a, \sigma) \quad \Leftrightarrow$$

$$\log k - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i + a)^2 \leq -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2 \quad \Leftrightarrow$$

$$2\sigma^2 \log k \leq \sum_{i=1}^n [(x_i + a)^2 - (x_i - a)^2] \quad \Leftrightarrow$$

$$\frac{\sigma^2}{2} \log k \leq a \sum_{i=1}^n x_i \quad \Leftrightarrow$$

$$\frac{\sigma^2}{2na} \log k \leq \bar{x} \quad (7.47)$$

Setzt man

$$c := \frac{\sigma^2}{2na} \log k \quad (7.48)$$

so entscheidet man für G sobald, $\bar{x} \geq c$.

Das k bzw. das c müssen nun noch nach den oben aufgestellten Kriterien bestimmt werden.

$$W = \{ \mathcal{U} = (x_1, \dots, x_n)^T \mid \bar{x} \geq \frac{\sigma^2}{2na} \log k \} \quad (7.49)$$

$$P(W) \leq \left(\frac{1}{k} \right)^{1-t} \exp \left\{ -\frac{2n}{\sigma^2} t(1-t) a^2 \right\} \quad (7.50)$$

$$1 - Q(W) \leq k^t \exp\left\{-\frac{2n}{\sigma^2} t(1-t) a^2\right\} \quad (7.51)$$

Löst man (7.48) nach k auf, so ergibt sich:

$$k = \exp \frac{2anc}{\sigma^2} \quad (7.52)$$

und

$$k^t = \exp \frac{2anct}{\sigma^2} \quad (7.53)$$

$$\left(\frac{1}{k}\right)^{1-t} = \exp - \frac{2anc(1-t)}{\sigma^2} \quad (7.54)$$

Somit gilt:

$$P(W) \leq \exp - \frac{2anc(1-t) + 2nt(1-t) a^2}{\sigma^2} \quad (7.55)$$

$$= \exp\left\{-\frac{2an(1-t)}{\sigma^2} [c + at]\right\} \quad (7.56)$$

Nun soll die Wahrscheinlichkeit für den Fehler 1. Art höchstens α betragen. Wir setzen daher:

$$\alpha = \exp\left\{-\frac{2an(1-t)}{\sigma^2} [c + at]\right\} \quad (7.57)$$

Daraus ergibt sich

$$c = \frac{\sigma^2 \log \alpha}{-2an(1-t)} - ta \quad (7.58)$$

Außerdem hat man

$$1 - Q(W) \leq \exp\left[\frac{2anct}{\sigma^2} - \frac{2n}{\sigma^2} t(1-t) a^2\right] \quad (7.59)$$

Setzt man für c ein, so wird

$$1 - Q(W) \leq \exp\left\{\frac{2ant}{\sigma^2} \left[\frac{\sigma^2 \log \alpha}{-2an(1-t)} - ta\right] - \frac{2n}{\sigma^2} t(1-t) a^2\right\} \quad (7.60)$$

und zusammengefaßt:

$$1 - Q(W) \leq \exp - \left(\frac{t \log \alpha}{1-t} + \frac{2nta^2}{\sigma^2}\right) \quad (7.61)$$

Die Wahrscheinlichkeit für den Fehler II. Art wird klein, wenn $k^t I(P_0, Q_0)(t)$ minimal wird.

Man minimiert $k^t I(P_0, Q_0)(t)$, wenn man

$$\bar{H} : [0,1] \rightarrow \mathbb{R}$$

$$t \rightarrow \bar{H}(t) = \frac{t \log \alpha}{1-t} + \frac{2nta^2}{\sigma^2} \quad (7.62)$$

maximiert.

Nun gilt

$$\frac{d\bar{H}(t)}{dt} = \log \alpha \frac{1}{(1-t)^2} + \frac{2na^2}{\sigma^2} \quad (7.63)$$

$$\frac{d^2\bar{H}(t)}{dt^2} = \log \alpha \frac{2}{(1-t)^3} \quad (7.64)$$

Da $\log \alpha < 0$, so folgt

$$\frac{d^2\bar{H}(t)}{dt^2} < 0 \text{ für alle } t \in [0,1] \quad (7.65)$$

Jeder kritische Wert von $\bar{H}(\cdot)$ ist somit ein Maximum. Zur Bestimmung der kritischen Werte wird

$$\frac{d\bar{H}(t)}{dt} = 0 \quad (7.66)$$

gesetzt. Man erhält

$$t_{\max} = 1 - \frac{\sigma}{a} \sqrt{\frac{\log 1/\alpha}{2n}} \quad (7.67)$$

Bemerkung:

Der Umfang n der Stichprobe muß dann immer so gewählt werden, daß

$$\left(1 - \frac{\sigma}{a} \sqrt{\frac{\log 1/\alpha}{2n}}\right) \in [0,1] \quad (7.68)$$

Das ist aber eine milde Forderung.¹⁾

Das zu $t_{\max}$ gehörende $c_{\max}$ berechnet sich zu

1) Diese Forderung kann man sogar fallenlassen, wenn man einen größeren Definitionsbereich von $I(P,Q)(\cdot)$ zuläßt. Im Falle der Normalverteilung existiert, wie sich zeigen läßt, $I(P,Q)(t)$ für alle $t \in \mathbb{R}$.

$$\begin{aligned}
 c_{\max} &= \frac{\sigma^2 \log \alpha}{-2an(1 - t_{\max})} - a \cdot t_{\max} \\
 &= \frac{\sigma^2 \log \alpha}{-2an\left(\frac{\sigma}{a} \sqrt{\frac{\log 1/\alpha}{2n}}\right)} - \left(1 - \frac{\sigma}{a} \sqrt{\frac{\log 1/\alpha}{2n}}\right)a \\
 &= -a + \sigma \sqrt{\frac{2}{n} \log 1/\alpha} \quad (7.69)
 \end{aligned}$$

Für die Wahrscheinlichkeit, den Fehler II. Art zu begehen, läßt sich vermöge $t_{\max}$ die obere Schranke angeben:

$$1 - Q(W) \leq \exp - \left(\frac{t_{\max} \log \alpha}{1 - t_{\max}} + \frac{2n t_{\max}^2 a^2}{\sigma^2} \right) \quad (7.70)$$

Nach einigen Umformungen erhält man

$$1 - Q(W) \leq \exp - \left(\sqrt{\log 1/\alpha} - \frac{a}{\sigma} \sqrt{2n} \right)^2 \quad (7.71)$$

Wir vergleichen jetzt diesen Test mit dem Test, der von Matusita, Suzuki und Hudimoto [45] vorgeschlagen wurde.

Bei diesen Autoren wurde aufgrund der Beziehungen

$$P(W) \leq \sqrt{\frac{1}{k}} I(P_o, Q_o) \left(\frac{1}{2} \right) \quad (7.72)$$

$$1 - Q(W) \leq \sqrt{k} I(P_o, Q_o) \left(\frac{1}{2} \right) \quad (7.73)$$

ähnlich wie oben vorgegangen, jedoch ergab sich dort

$$c = -\frac{a}{2} + \frac{\log 1/\alpha}{na} \quad (7.74)$$

Als obere Schranke für die Wahrscheinlichkeit für den Fehler II. Art erhielt man dort

$$1 - Q(W) \leq \exp \left[-na^2 + \log 1/\alpha \right] \quad (7.75)$$

(bei $\sigma = 1$)

Bei unseren Untersuchungen hatte sich ergeben

$$1 - Q(W) \leq \exp - (\sqrt{\log 1/\alpha} - a\sqrt{2n})^2 \quad (7.76)$$

(bei $\sigma = 1$)

Da aber

$$-a^2 n + \log 1/\alpha + (\sqrt{\log 1/\alpha} - a\sqrt{2n})^2 = (\sqrt{2 \log 1/\alpha} - \sqrt{na})^2 \geq 0 \quad (7.77)$$

geht bei dem hier vorliegenden Test zum Niveau α bei wachsendem n die Schranke für die Wahrscheinlichkeit, den Fehler II. Art zu begehen, viel rascher gegen Null als jene beim Test der Japaner.

Wie man sofort findet, gilt (bei $\sigma = 1$):

$$c - c_{\max} = \frac{a}{2} + \frac{\sqrt{\log 1/\alpha}}{a} \left[\frac{\log 1/\alpha}{n} - a \sqrt{\frac{2}{n}} \right] \quad (7.77 \text{ a})$$

Für genügend großes n gilt mithin

$$c_{\max} < c \quad (7.77 \text{ b})$$

Daraus erkennt man, daß bei dem hier vorgeschlagenen Test auch die Wahrscheinlichkeit für den Fehler II. Art schneller gegen Null konvergiert als die entsprechende Wahrscheinlichkeit beim Test der Japaner. Übrigens kann man ja auch $\text{Prob}(\bar{X} \leq c_{\max} | H_1) = \beta$ direkt berechnen (bei $\{P_0\}$ gegen $\{Q_0\}$).

Man erhält

$$\beta = F_Z(-2a\sqrt{n} + \sqrt{2 \log 1/\alpha}) \quad (7.77 \text{ c})$$

wobei F_Z die Verteilungsfunktion einer Zufallsvariablen Z mit $Z \sim N(0,1)$ ist.

Für das Testproblem H gegen G gilt also, wie man sich dann überlegt:

$$\beta \leq F_Z(-2a\sqrt{n} + \sqrt{2 \log 1/\alpha}) \quad (7.77 \text{ d})$$

Beispiel 2:

$(X_1, \dots, X_n)$ sei die Stichprobe einer normalverteilten Grundgesamtheit mit unbekanntem μ und bekanntem σ (Stichprobenzüge unabhängig). Das Testproblem möge lauten:

$$H : |\mu| \leq a_1 \quad (7.78)$$

$$G : |\mu| \geq a_2 \quad (7.79)$$

wobei $a := a_2 - a_1 > 0$, $a_1 > 0$.

Mit den Bezeichnungen wie in vorigem Beispiel gilt:

Trifft H zu, so ist

$$f(x_1, \dots, x_n | \mu_1, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2\right\}, |\mu_1| \leq a_1 \quad (7.80)$$

Trifft G zu, so ist

$$g(x_1, \dots, x_n | \mu_2, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_2)^2\right\}, |\mu_2| \geq a_2 \quad (7.81)$$

$$I(P, Q)(t) = \exp\left\{-\frac{n}{2\sigma^2} t(1-t)(\mu_1 - \mu_2)^2\right\} \quad (7.82)$$

wird bei festem $t \in [0, 1)$ maximal, wenn

$$|\mu_2 - \mu_1| = a \quad (7.83)$$

Das Maximum wird in zwei Fällen angenommen:

$$\text{Fall I: } \mu_1 = -a_1, \quad \mu_2 = -a_2 \quad (7.84)$$

$$\text{Fall II: } \mu_1 = a_1, \quad \mu_2 = a_2 \quad (7.85)$$

Wir testen zunächst im I. Fall die einfachen Hypothesen

$$H_1 : \mu = -a_1 \quad (7.86)$$

$$G_1 : \mu = -a_2 \quad (7.87)$$

und dann im II. Fall die einfachen Hypothesen

$$H_2 : \mu = a_1 \quad (7.88)$$

$$G_2 : \mu = a_2 \quad (7.89)$$

I. Fall

$$I(P_0, Q_0)(t) = \exp\left\{-\frac{n}{2\sigma^2} t(1-t)a^2\right\} \quad (7.90)$$

Bestimmung von k:

$$k = f(\mathcal{U} | -a_1, \sigma) \leq g(\mathcal{U} | -a_2, \sigma) \Leftrightarrow$$

$$\log k - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i + a_1)^2 \leq -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i + a_2)^2 \Leftrightarrow$$

$$2\sigma^2 \log k \leq \sum_{i=1}^n (2a_1 x_i + a_1^2 - 2a_2 x_i - a_2^2) \quad \Leftarrow$$

$$\bar{x} \leq - \frac{\sigma^2 \log k}{na} - \frac{a_1 + a_2}{2} \quad (7.91)$$

Das kritische Gebiet W_1 für diesen Test lautet

$$W_1 = \{(x_1, \dots, x_n) | \bar{x} \leq - \left(\frac{\sigma^2 \log k}{na} + \frac{a_1 + a_2}{2} \right)\} \quad (7.92)$$

II. Fall

Hier zeigt sich analog zum I. Fall für das kritische Gebiet

$$W_2 = \{(x_1, \dots, x_n) | \bar{x} \geq \left(\frac{\sigma^2 \log k}{na} + \frac{a_1 + a_2}{2} \right)\} \quad (7.93)$$

und

$$I(P_0, Q_0)(t) = \exp\left\{-\frac{n}{2\sigma^2} t(1-t)a^2\right\} \quad (7.94)$$

Zum Testen von H gegen G wählt man dann als kritisches Gebiet (nicht gleichmäßig bestes! - vgl. z.B. Witting [56] S. 41):

$$W : = W_1 \cup W_2 \quad (7.95)$$

Setzt man

$$c_1 : = - \left(\frac{\sigma^2 \log k}{na} + \frac{a_1 + a_2}{2} \right) \quad (7.96)$$

$$c_2 : = -c_1 \quad (7.97)$$

so entscheidet man für H , falls

$$\bar{x} \in (c_1, c_2) \quad (7.98)$$

und für G , falls

$$\bar{x} \leq c_1 \quad \text{oder} \quad \bar{x} \geq c_2 \quad (7.99)$$

Löst man nach k auf, so ergibt sich

$$k = \exp\left\{\frac{na}{\sigma^2} \left(c_2 - \frac{a_1 + a_2}{2}\right)\right\} \quad (7.100)$$

sowie

$$\left(\frac{1}{k}\right)^{1-t} = \exp\left\{-\frac{na}{\sigma^2} \left(c_2 - \frac{a_1 + a_2}{2}\right)(1-t)\right\} \quad (7.101)$$

$$k^t = \exp\left\{\frac{na}{\sigma^2} \left(c_2 - \frac{a_1 + a_2}{2}\right)t\right\} \quad (7.102)$$

Da

$$P(W) = P(W_1) + P(W_2) \quad (7.103)$$

gilt für P

$$P(W) \leq 2\left(\frac{1}{k}\right)^{1-t} I(P_O, Q_O)(t) \quad (7.104)$$

Nun ist dann

$$\left(\frac{1}{k}\right)^{1-t} I(P_O, Q_O)(t) = \exp\left\{-\frac{na}{\sigma^2} (1-t)\left(c_2 - \frac{a_1 + a_2}{2} + \frac{ta}{2}\right)\right\} \quad (7.105)$$

Um einen Test zum Niveau α zu bekommen, setze man

$$\alpha = 2\left(\frac{1}{k}\right)^{1-t} I(P_O, Q_O)(t) \quad (7.106)$$

Daraus folgt

$$c_2 = -\frac{\sigma^2 \log \frac{\alpha}{2}}{na(1-t)} + \frac{a_1 + a_2}{2} - \frac{ta}{2} \quad (7.107)$$

Da für $Q \in \mathcal{Q}$

$$Q(W_j) \geq 1 - k^t I(P_O, Q_O)(t) \quad j = 1, 2 \quad (7.108)$$

so folgt:

$$1 - Q(W) \leq 2 k^t I(P_O, Q_O)(t) - 1 \quad (7.109)$$

d.h.

$$1 - Q(W) \leq 2 \exp\left\{\frac{na}{\sigma^2} t\left(c_2 - \frac{a_1 + a_2}{2} - \frac{(1-t)a}{2}\right)\right\} - 1 \quad (7.110)$$

$2 k^t I(P_O, Q_O)(t) - 1$ wird minimal, wenn man die Funktion

$$\bar{H} : [0, 1) \rightarrow \mathbb{R}$$

$$t \rightarrow \bar{H}(t) = -\frac{na}{\sigma^2} t\left(c_2 - \frac{a_1 + a_2}{2} - \frac{(1-t)a}{2}\right) \quad (7.111)$$

maximiert. c_2 hängt allerdings noch von t ab. Setzt man ein, so ist

$$\bar{H}(t) = \frac{t}{1-t} \log \frac{\alpha}{2} + \frac{na^2 t}{2\sigma^2} \quad (7.112)$$

woraus

$$\frac{d\bar{H}(t)}{dt} = \frac{\log \frac{\alpha}{2}}{(1-t)^2} + \frac{na^2}{2\sigma^2} \quad (7.113)$$

$$\frac{d^2\bar{H}(t)}{dt^2} = \frac{2 \log \frac{\alpha}{2}}{(1-t)^3} < 0 \quad \text{für alle } t \in [0,1) \quad (7.114)$$

folgt.

Jeder kritische Wert von $\bar{H}$ ist somit ein Maximum. Zur Bestimmung derselben setzen wir

$$\frac{d\bar{H}(t)}{dt} = 0 \quad (7.115)$$

Man erhält

$$t_{\max} = 1 - \frac{\sigma}{a} \sqrt{\frac{2}{n} \log \frac{2}{\alpha}} \quad (7.116)$$

und für den Wert c_2 ergibt sich

$$c_{2\max} = a_1 + \sigma \sqrt{\frac{2}{n} \log \frac{2}{\alpha}} \quad (7.117)$$

wobei n wieder so groß gewählt sei, daß $t_{\max} \in [0,1)$.

Es ist weiter

$$c_{1\max} = -a_1 - \sigma \sqrt{\frac{2}{n} \log \frac{2}{\alpha}} \quad (7.118)$$

Eine obere Schranke für die Wahrscheinlichkeit, den Fehler 2. Art zu begehen, erhält man aus der Relation

$$1 - Q(W) \leq 2 \exp \left\{ \frac{na}{\sigma^2} t_{\max} \left(c_{2\max} - \frac{a_1 + a_2}{2} - \frac{(1 - t_{\max})a}{2} \right) \right\} - 1 \quad (7.119)$$

oder eingesetzt

$$1 - Q(W) \leq 2 \exp \left\{ - \frac{n}{2\sigma^2} \left(a - \sigma \sqrt{\frac{2}{n} \log \frac{2}{\alpha}} \right)^2 \right\} - 1 \quad (7.120)$$

Auch hier läßt sich zeigen, daß dieser Test "besser" ist als ein ähnlicher Test der oben erwähnten Japaner.

Beispiel 3:

$(X_1, \dots, X_n)$ sei die Stichprobe einer normalverteilten Grundgesamtheit mit unbekanntem μ und bekanntem σ . (Unabhängige Stichprobenzüge.)

Wir betrachten das einfache Testproblem:

$$H : \mu = \mu_1 \quad (7.121)$$

$$G : \mu = \mu_2 \quad (7.122)$$

wobei ohne Einschränkung $\mu_1 < \mu_2$ sei. (Der Fall $\mu_1 > \mu_2$ verläuft analog.)

Die Bezeichnungen seien wie in den anderen Beispielen gewählt.

Trifft H zu, so ist

$$f(x_1, \dots, x_n | \mu_1, \sigma) = \frac{1}{\sqrt{2\pi} \sigma^2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2\right\} \quad (7.123)$$

Trifft G zu, so ist

$$g(x_1, \dots, x_n | \mu_2, \sigma) = \frac{1}{\sqrt{2\pi} \sigma^2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_2)^2\right\} \quad (7.124)$$

Bei festem $t \in [0, 1)$ ist in diesem Fall

$$d_t(\mathcal{P}, \mathcal{Q}) = I(P_0, Q_0)(t) \quad (7.125)$$

mit

$$I(P_0, Q_0)(t) = \exp\left\{-\frac{n}{2\sigma^2} t(1-t)(\mu_1 - \mu_2)^2\right\} \quad (7.126)$$

Zur Bestimmung des kritischen Gebiets betrachte man die Ungleichung

$$k f(x_1, \dots, x_n | \mu_1, \sigma) \leq g(x_1, \dots, x_n | \mu_2, \sigma) \quad (7.127)$$

Durch einfache Umformungen sieht man dann, daß (7.127) äquivalent ist mit

$$\bar{x} \geq \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2 \log k}{n(\mu_2 - \mu_1)} \quad (7.128)$$

Setzt man

$$c := \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2 \log k}{n(\mu_2 - \mu_1)} \quad (7.129)$$

so folgt:

$$k = \exp - \frac{n(\mu_2^2 - \mu_1^2) - 2nc(\mu_2 - \mu_1)}{2\sigma^2}$$

und damit

$$\left(\frac{1}{k}\right)^{1-t} = \exp\left\{\frac{n(\mu_2^2 - \mu_1^2) - 2nc(\mu_2 - \mu_1)}{2\sigma^2} (1-t)\right\} \quad (7.131)$$

$$k^t = \exp\left\{-\frac{n(\mu_2^2 - \mu_1^2) - 2nc(\mu_2 - \mu_1)}{2\sigma^2} t\right\} \quad (7.132)$$

Wird

$$W = \{(x_1, \dots, x_n) | \bar{x} \geq c\} \quad (7.133)$$

gesetzt, so gilt für P

$$\begin{aligned} P(W) &\leq \left(\frac{1}{k}\right)^{1-t} I(P_0, Q_0)(t) \\ &= \exp\left\{-\frac{n(1-t)(\mu_2 - \mu_1)}{2\sigma^2} (-\mu_2 - \mu_1 + 2c + t(\mu_2 - \mu_1))\right\} \end{aligned} \quad (7.134)$$

Um die Wahrscheinlichkeit für den Fehler 1. Art kleiner als α zu machen, setzen wir

$$\left(\frac{1}{k}\right)^{1-t} I(P_0, Q_0)(t) = \alpha \quad (7.135)$$

Dann ist nach Auflösen

$$c = \frac{1}{2} [\mu_1 + \mu_2 - t(\mu_2 - \mu_1) - \frac{2\sigma^2 \log \alpha}{n(1-t)(\mu_2 - \mu_1)}] \quad (7.136)$$

Da für Q

$$1 - Q(W) \leq k^t I(P_0, Q_0)(t) \quad (7.137)$$

so ist nach Einsetzen von c

$$1 - Q(W) \leq \exp\left\{\frac{-n(\mu_2 - \mu_1)^2 t}{2\sigma^2} - \log \alpha \frac{t}{1-t}\right\} \quad (7.138)$$

Um die Wahrscheinlichkeit für den Fehler 2. Art zu minimieren, maximiert man die Funktion

$$\bar{H} : [0,1) \rightarrow \mathbb{R}$$

$$t \rightarrow \bar{H}(t) = \frac{n(\mu_2 - \mu_1)^2 t}{2\sigma^2} + \log \alpha \frac{t}{1-t} \quad (7.139)$$

Es ist

$$\frac{d\bar{H}(t)}{dt} = \frac{n(\mu_2 - \mu_1)^2}{2\sigma^2} + \log \alpha \frac{t}{(1-t)^2} \quad (7.140)$$

$$\frac{d^2\bar{H}(t)}{dt^2} = 2 \log \alpha \frac{1}{(1-t)^3} \quad (7.141)$$

Weil $\log \alpha < 0$, ist für alle $t \in [0,1)$

$$\frac{d^2\bar{H}(t)}{dt^2} < 0 \quad (7.142)$$

Durch Nullsetzung der 1. Ableitung $\frac{d\bar{H}(t)}{dt}$ erhält man einen kritischen Wert, der dann ein Maximum von $\bar{H}$ sein muß. Es folgt

$$t_{\max} = 1 - \sigma \sqrt{\frac{2 \log 1/\alpha}{n(\mu_2 - \mu_1)^2}} \quad (7.143)$$

Auch hier gilt für genügend großes n :

$$t_{\max} \in [0, 1) \quad (7.144)$$

Der Wert $c_{\max}$ berechnet sich zu

$$c_{\max} = \mu_1 + \sigma \sqrt{\frac{2}{n} \log 1/\alpha} \quad (7.145)$$

Für die Wahrscheinlichkeit, den Fehler II. Art zu begehen, ergibt sich nach einfachen Umformungen die Beziehung

$$1 - Q(W) \leq \exp\left\{-\frac{n}{2\sigma^2} (\mu_2 - \mu_1 - \sigma \sqrt{\frac{2}{n} \log 1/\alpha})^2\right\} \quad (7.146)$$

Als kleine Ergänzung sei vermerkt, daß

$$1 - Q(W) \leq \alpha \quad (7.147)$$

sofern

$$n \geq \frac{8\sigma^2}{(\mu_2 - \mu_1)^2} \log 1/\alpha \quad (7.148)$$

Bemerkung 1:

Zu dem vorliegenden Problem gibt es bekanntlich einen sogenannten gleichmäßig besten Test, der die vorgegebene Irrtumswahrscheinlichkeit α voll ausnutzt. Als kritisches Gebiet wird dabei

$$W' = \{(x_1, \dots, x_n) | \bar{x} \geq c'\} \quad (7.149)$$

mit

$$c' = \mu_1 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \quad (7.150)$$

wobei

$$\int_{-\infty}^{z_{1-\alpha}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} x^2\right\} dx = 1 - \alpha \quad (7.151)$$

so daß dann

$$P(W') = \alpha \quad (7.152)$$

Vergleicht man c' mit $c_{\max}$, so erkennt man, daß stets

$$c_{\max} \geq c' \quad (7.153)$$

(Man beachte, daß W so gewählt wurde, daß

$P(W) \leq \left(\frac{1}{k}\right)^{1-t} I(P, Q)(t) = \alpha$). Im allgemeinen wird sich also ergeben, daß

$$P(W) < \alpha \quad (7.154)$$

Die folgende Tabelle, die unter Verwendung des Buches von Wetzel, W.; Jöhnk, M.D.; Naeve, P. [54] erstellt wurde, zeigt die Relation zwischen dem vorgegebenen α und der tatsächlichen Wahrscheinlichkeit, den Fehler 1. Art zu begehen, wenn man den Test aus Beispiel 3 anwendet. Dabei entspricht $\sqrt{2 \log 1/\alpha}$ dem Wert $z_{1-\alpha}$ im klassischen Test. Z ist eine normalverteilte Zufallsvariable mit $E(Z) = 0$, $\text{Var}(Z) = 1$.

α	$\sqrt{2 \log 1/\alpha}$	$\text{Prob}(Z \geq \sqrt{2 \log 1/\alpha})$
0.001	3.7169	0.0001
0.01	3.0349	0.0012
0.03	2.6482	0.0041
0.05	2.4478	0.0072
0.0084	3.0902	0.001
0.0668	2.3263	0.01
0.1706	1.8808	0.03
0.2585	1.6449	0.05

Wie man sieht, kann man durch anfängliche Wahl eines α^* größer als α erreichen, daß

$$P(W) = \alpha \quad (7.155)$$

jedoch würde dieses Vorgehen das Vorliegen einer zusätzlichen Tabelle erfordern.

Bemerkung 2:

Bei allen obigen drei Beispielen ist noch aufgefallen, daß die kritischen Bereiche sehr schnell ohne irgendwelche Tabellen (nur evtl. mit Hilfe eines Rechenschiebers) angegeben werden können. Zudem kann bequem eine obere Schranke für die Wahrscheinlichkeit des Fehlers 2. Art berechnet werden. Insgesamt gesehen sind diese Tests also sehr handlich. Ein Nachteil liegt darin, daß das gegebene α nicht vollständig ausgenützt wird.

Im letzten Teil dieses Kapitels werden Ergebnisse von Kraft [38] verallgemeinert. Die von dem Autor stammenden Beweisideen werden dazu leicht modifiziert.

$(X_i, \mathcal{X}_i)_{i=1}^{\infty}$ sei eine Folge von Meßräumen. Der Kürze halber setze man

$$X^{(n)} : = \prod_{i=1}^n X_i \quad X : = \prod_{i=1}^{\infty} X_i \quad (7.156)$$

$$\mathcal{X}^{(n)} : = \bigotimes_{i=1}^n \mathcal{X}_i \quad \mathcal{X} : = \bigotimes_{i=1}^{\infty} \mathcal{X}_i \quad (7.157)$$

Die σ -Algebra $\mathcal{X}^{(n)}$ und die σ -Algebra

$$\sigma\left(\left\{E \times \prod_{i=n+1}^{\infty} X_i \mid E \in \mathcal{X}^{(n)}\right\}\right) \quad (7.158)$$

werden identifiziert. Entsprechend wird $\mathcal{X}^{(n)}$ in $\mathcal{X}$ eingebettet.

$\{P^{(n)}, Q^{(n)}\}_{n=1}^{\infty}$ sei eine Doppelfolge von Wahrscheinlichkeitsmaßen, jeweils $P^{(n)}$ bzw. $Q^{(n)}$ auf $\mathcal{X}^{(n)}$, die sich eindeutig auf $\mathcal{X}$ zu den Wahrscheinlichkeitsmaßen P bzw. Q fortsetzen lassen. Insbesondere gilt dann für $E \in \mathcal{X}^{(n)}$:

$$P^{(n)}(E) = P\left(E \times \prod_{i=n+1}^{\infty} X_i\right) \quad (7.159)$$

$$Q^{(n)}(E) = Q\left(E \times \prod_{i=n+1}^{\infty} X_i\right) \quad (7.160)$$

SATZ 7.4

Für alle $t \in [0,1]$ gilt:

$$I(P^{(n)}, Q^{(n)})(t) \leq I(P^{(n-1)}, Q^{(n-1)})(t) \quad (7.161)$$

Beweis:

Folgt unmittelbar aus Korollar 1.9 und der Tatsache, daß:

$$P^{(n)} / \mathcal{X}^{(n-1)} = P^{(n-1)} \quad (7.162)$$

$$Q^{(n)} / \mathcal{X}^{(n-1)} = Q^{(n-1)} \quad (7.163)$$

Wir beschäftigen uns nun mit dem Problem des Auffindens konsistenter Tests (siehe auch Schmetterer [51] S. 290). Dazu zunächst die

Definition:

Zu testen seien die Hypothesen H und G . (H sei durch die Familie $\mathcal{P}$ von Wahrscheinlichkeitsmaßen, G durch die Familie $\mathcal{Q}$ von Wahrscheinlichkeitsmaßen charakterisiert.)

Eine Folge

$$\phi_n : X \rightarrow [0,1]$$

($\mathcal{X}^{(n)} - \mathcal{Q}^1 \cap [0,1]$) - meßbarer Funktionen heißt konsistenter Test zum Testproblem H gegen G , wenn gilt:

$$\bigwedge_{P \in \mathcal{P}} \lim_{n \rightarrow \infty} \int \phi_n dP = 0 \quad (7.164)$$

$$\bigwedge_{Q \in \mathcal{Q}} \lim_{n \rightarrow \infty} \int \phi_n dQ = 1 \quad (7.165)$$

SATZ 7.5

Die folgenden Aussagen sind äquivalent:

$$(i) \quad P \perp Q \quad (7.166)$$

$$(ii) \quad \lim_{n \rightarrow \infty} I(P^{(n)}, Q^{(n)})(t_0) = 0 \text{ für ein } t_0 \in (0,1) \quad (7.167)$$

$$(iii) \quad \text{Es gibt einen konsistenten Test zum Problem } H \text{ gegen } G \text{ mit} \\ \mathcal{P} = \{P\}, \quad \mathcal{Q} = \{Q\}.$$

Beweis:

$$(i) \Rightarrow (ii)$$

Sei $t_0 \in (0,1)$ beliebig und

$$\lambda_n := \frac{P^{(n)} + Q^{(n)}}{2} \quad (7.168)$$

$$\lambda := \frac{P + Q}{2} \quad (7.169)$$

Dann gilt

$$f_n = E_\lambda(f | \mathcal{X}^{(n)}) \quad [\lambda] \quad (7.170)$$

$$g_n = E_\lambda(g | \mathcal{X}^{(n)}) \quad [\lambda] \quad (7.171)$$

mit

$$f_n = \frac{dP^{(n)}}{d\lambda} \quad (7.172)$$

$$g_n = \frac{dQ^{(n)}}{d\lambda} \quad (7.173)$$

Wie man sieht, sind

$$\{f_n, \mathcal{X}^{(n)}\}, \quad \{g_n, \mathcal{X}^{(n)}\} \quad (7.174)$$

Martingale, und es gilt weiter

$$\lim_{n \rightarrow \infty} f_n = f \quad [\lambda] \quad (7.175)$$

$$\lim_{n \rightarrow \infty} g_n = g \quad [\lambda] \quad (7.176)$$

Dann ist aber auch

$$\lim_{n \rightarrow \infty} f_n^{t_0} g_n^{1-t_0} = f^{t_0} g^{1-t_0} \quad [\lambda] \quad (7.177)$$

Nach dem Satz von der majorisierten Konvergenz von Lebesgue gilt mithin:

$$\lim_{n \rightarrow \infty} I(P^{(n)}, Q^{(n)})(t_0) = I(P, Q)(t_0) \quad (7.178)$$

Nach Voraussetzung war $I(P, Q)(t_0) = 0$, d.h. (ii) ist erfüllt.

(ii) $\Rightarrow$ (iii)

Es gebe ein $t_0 \in (0, 1)$ mit

$$\lim_{n \rightarrow \infty} I(P^{(n)}, Q^{(n)})(t_0) = 0 \quad (7.179)$$

Wir setzen

$$E_n := \{(x_1, \dots, x_n) \mid (x_1, \dots, x_n) \in \prod_{i=1}^n X_i \wedge f_n(x_1, \dots, x_n) \leq g_n(x_1, \dots, x_n)\} \quad (7.180)$$

Wähle

$$\phi_n := \chi_{E_n} \quad (7.181)$$

Es ist

$$\begin{aligned} \int \phi_n dP &= P(E_n) \\ &= \int_{E_n} f_n d\lambda \\ &= \int_{E_n} f_n^{t_0} f_n^{1-t_0} d\lambda \\ &\leq I_{E_n}(P^{(n)}, Q^{(n)})(t_0) \\ &\leq I(P^{(n)}, Q^{(n)})(t_0) \end{aligned} \quad (7.182)$$

und weiter

$$\begin{aligned} \int \phi_n dQ &= Q(E_n) \\ &= \int_{E_n} g_n d\lambda \end{aligned}$$

$$\begin{aligned}
 &= 1 - \int \frac{t_0^{1-t_0}}{\bar{E}_n} \cdot \bar{E}_n \\
 &\geq 1 - \int \frac{f_n^{t_0} \bar{E}_n^{1-t_0}}{\bar{E}_n} d\lambda \\
 &= 1 - I_{\bar{E}_n} (P^{(n)}, Q^{(n)})(t_0) \\
 &\geq 1 - I(P^{(n)}, Q^{(n)})(t_0) \quad (7.183)
 \end{aligned}$$

Wegen der Voraussetzung folgt

$$\lim_{n \rightarrow \infty} \int \phi_n dP = 0 \quad (7.184)$$

$$\lim_{n \rightarrow \infty} \int \phi_n dQ = 1 \quad (7.185)$$

(iii) $\Rightarrow$ (i)

Es existiere ein konsistenter Test $\{\phi_n\}_{n \in \mathbb{N}}$ zum Testproblem H gegen G .
 Man wähle eine Folge $\{\epsilon_k\}_{k=1}^\infty$, $\epsilon_k > 0$ und eine Teilfolge $\{n_k\}_{k=1}^\infty$
 der natürlichen Zahlen mit

$$\int \phi_{n_k} f d\lambda < \epsilon_k \quad (7.186)$$

$$\int \phi_{n_k} g d\lambda > 1 - \epsilon_k \quad (7.187)$$

$$\sum_{k=1}^\infty \epsilon_k < \infty \quad (7.188)$$

Setzt man

$$\phi := \limsup_{k \rightarrow \infty} \phi_{n_k} \quad (7.189)$$

so folgt

$$\begin{aligned}
 \int \phi f d\lambda &= \lim_{j \rightarrow \infty} \int \left(\sup_{k \geq j} \phi_{n_k} \right) f d\lambda \\
 &\leq \lim_{j \rightarrow \infty} \sum_{k \geq j} \int \phi_{n_k} f d\lambda
 \end{aligned}$$

$$\leq \lim_{j \rightarrow \infty} \sum_{k \geq j} \epsilon_k$$

$$= 0 \quad (7.190)$$

Es ist also

$$P(\phi = 0) = 1 \quad (7.191)$$

Analog zeigt man

$$Q(\phi = 0) = 0 \quad (7.192)$$

Dies ist nichts anderes als

$$P \perp Q \quad (7.193)$$

Beispiel:

(Man vergleiche Kraft [38] S. 219.)

$P^{(n)}$ sei das Bildmaß bez. der vektorwertigen Zufallsvariablen $\vec{X}$ mit der Normalverteilung

$$N(\vec{\mu}_n, V_n) \quad (7.194)$$

und $Q^{(n)}$ sei das Bildmaß bez. der vektorwertigen Zufallsvariablen $\vec{X}$ mit der Normalverteilung

$$N(\vec{\rho}_n, V_n) \quad (7.195)$$

Es gelte

$$\lim_{n \rightarrow \infty} (\vec{\mu}_n - \vec{\rho}_n)^T V_n^{-1} (\vec{\mu}_n - \vec{\rho}_n) = \infty \quad (7.196)$$

(etwa wenn

$$\vec{\mu}_n = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \vec{\rho}_n = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad V_n = \begin{pmatrix} 1/n & 0 & 0 & 0 \\ 0 & 1/n & 0 & 0 \\ 0 & 0 & 1/n & 0 \\ 0 & 0 & 0 & 1/n \end{pmatrix} \quad (7.197)$$

Dann existiert ein konsistenter Test $\{\phi_n\}_{n \in \mathbb{N}}$ zum Testproblem $\{P\}$ gegen $\{Q\}$, wobei P und Q wie auf S. 93 definiert seien.

Es ist nämlich (vgl. Anhang) für $t_0 \in (0,1)$

$$I(P^{(n)}, Q^{(n)})(t_0) = \exp\left[-\frac{t_0}{2} (1-t_0) (\vec{\mu}_n - \vec{\rho}_n)^T V_n^{-1} (\vec{\mu}_n - \vec{\rho}_n)\right] \quad (7.198)$$

und damit

$$\lim_{n \rightarrow \infty} I(P^{(n)}, Q^{(n)})(t_0) = 0 \quad (7.199)$$

Der Rest folgt aus Satz 7.5.

Anhang

Einige wichtige AEF

Diskrete Verteilungen

Verteilung	Dichte von P	Dichte von Q	Relation der Parameter	$I(P, Q)(t)$
Zweipunkt- verteilung	$P(x_1) = p$ $P(x_2) = 1 - p$	$Q(x_1) = q$ $Q(x_2) = 1 - q$	$0 < p < 1$ $0 < q < 1$	$p^t q^{1-t} + (1-p)^t (1-q)^{1-t}$
Binomial- verteilung	$\binom{n}{x} p^x (1-p)^{n-x}$ $x \in \mathbb{N} \cup \{0\}$	$\binom{n}{x} q^x (1-q)^{n-x}$ $x \in \mathbb{N} \cup \{0\}$	$0 < p < 1$ $0 < q < 1$	$[p^t q^{1-t} + (1-p)^t (1-q)^{1-t}]^n$
Poisson- verteilung	$\frac{\lambda_1^x}{x!} \exp(-\lambda_1)$ $x \in \mathbb{N} \cup \{0\}$	$\frac{\lambda_2^x}{x!} \exp(-\lambda_2)$ $x \in \mathbb{N} \cup \{0\}$	$\lambda_i > 0$ $i = 1, 2$	$\exp\{-\lambda_1 t - \lambda_2(1-t) + \lambda_1^t \lambda_2^{1-t}\}$
Pascal- verteilung	$(1-p)^x p^x$ $x \in \mathbb{N} \cup \{0\}$	$(1-q)^x q^x$ $x \in \mathbb{N} \cup \{0\}$	$0 < p < 1$ $0 < q < 1$	$\frac{(1-p)^t (1-q)^{1-t}}{1 - p^t q^{1-t}}$

Stetige Verteilungen

Verteilung	Dichte von P	Dichte von Q	Relation der Parameter	I(P, Q)(t)
Rechteckverteilung	$f(x) = \begin{cases} \frac{1}{v_1} & 0 \leq x < v_1 \\ 0 & \text{sonst} \end{cases}$	$g(x) = \begin{cases} \frac{1}{v_2} & 0 \leq x < v_2 \\ 0 & \text{sonst} \end{cases}$	$0 < v_1 < v_2$	$\left(\frac{v_1}{v_2} \right)^{1-t}$
Betaverteilung	$f(x) = \begin{cases} \frac{1}{B(p_1, q_1)} x^{p_1-1} (1-x)^{q_1-1} & \text{für } x \in [0, 1] \\ 0 & \text{sonst} \end{cases}$	$g(x) = \begin{cases} \frac{1}{B(p_2, q_2)} x^{p_2-1} (1-x)^{q_2-1} & \text{für } x \in [0, 1] \\ 0 & \text{sonst} \end{cases}$	$p_i > 0$ $q_i > 0$ $i = 1, 2$	$\frac{B(p_1 t + p_2(1-t), q_1 t + q_2(1-t))}{B(p_1, q_1)^t B(p_2, q_2)^{1-t}}$

Verteilung	Dichte von P	Dichte von Q	Relation der Parameter	$I(P, Q)(t)$
Normal- verteilung				$\text{Sonderfall: } \mu_1 = \mu_2 = \mu, \quad \frac{\sigma_1^t \sigma_2^{1-t}}{\sqrt{(1-t)\sigma_1^2 + t\sigma_2^2}}$
	$f(x) = \frac{1}{\sqrt{2\pi} \sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_1}\right)^2\right\}$	$g(x) = \frac{1}{\sqrt{2\pi} \sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_2}\right)^2\right\}$	$\sigma_i > 0$ $i = 1, 2$	$\text{Sonderfall: } \sigma_1 = \sigma_2 = \sigma, \quad \exp\left\{-\frac{1}{2\sigma^2} t(1-t)(\mu_1 - \mu_2)^2\right\}$
				$\text{Allgemeiner Fall: } \frac{\sigma_1^{1-t} \sigma_2^t}{\sqrt{(1-t)\sigma_1^2 + \sigma_2^2 t}} \exp\left\{-\frac{1}{2(\sigma_1 \sigma_2)^2} w\right\} \text{ mit}$ $w = (\mu_1 \sigma_2)^2 t + (\mu_2 \sigma_1)^2 (1-t) - \frac{[\sigma_2^2 \mu_1 t + \sigma_1^2 \mu_2 (1-t)]^2}{t\sigma_2^2 + (1-t)\sigma_1^2}$

Verteilung	Dichte von P	Dichte von Q	Relation der Parameter	$I(P, Q)(t)$
Multi-normal-verteilung	$f(\vec{x}_1, \dots, \vec{x}_n) =$ $\frac{\sqrt{\det R}}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})'R(\vec{x}-\vec{\mu})\right\}$	$g(\vec{x}_1, \dots, \vec{x}_n) =$ $\frac{\sqrt{\det R}}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{p})'R(\vec{x}-\vec{p})\right\}$	R positiv definit	$\exp\left\{-\frac{1}{2}t(1-t)(\vec{\mu}-\vec{p})'R(\vec{\mu}-\vec{p})\right\}$
Pareto-verteilung	$f(x) = \begin{cases} (v_1-1)x^{-v_1} & x \in [1, \infty) \\ 0 & \text{sonst} \end{cases}$	$g(x) = \begin{cases} (v_2-1)x^{-v_2} & x \in [1, \infty) \\ 0 & \text{sonst} \end{cases}$	$v_i > 0$ $i = 1, 2$	$\frac{(v_1-1)^t (v_2-1)^{1-t}}{v_1 t + v_2(1-t) - 1}$
Gamma-verteilung	$f(x) = \begin{cases} \frac{b_1^{p_1}}{\Gamma(p_1)} x^{p_1-1} e^{-b_1 x} & x \geq 0 \\ 0 & \text{sonst} \end{cases}$	$g(x) = \begin{cases} \frac{b_2^{p_2}}{\Gamma(p_2)} x^{p_2-1} e^{-b_2 x} & x \geq 0 \\ 0 & \text{sonst} \end{cases}$	$b_i > 0$ $p_i > 0$ $i = 1, 2$	$\frac{\left[\frac{b_1^{p_1}}{\Gamma(p_1)}\right]^t \left[\frac{b_2^{p_2}}{\Gamma(p_2)}\right]^{1-t}}{\left[b_1 t + b_2(1-t)\right]^{p_1 t + p_2(1-t)}}$

Verteilung	Dichte von P	Dichte von Q	Relation der Parameter	$I(P, Q)(t)$
χ^2 - verteilung *			$b_1 = b_2 = \frac{1}{2}$ $p_1 = \frac{n_1}{2}$ $p_2 = \frac{n_2}{2}$	$\frac{\left[\left(\frac{1}{2} (n_1 t + n_2 (1-t)) \right) \right]}{\left[\left(\frac{n_1}{2} \right)^t \left[\left(\frac{n_2}{2} \right)^{1-t} \right] \right]}$
Exponential- verteilung *			$b_i > 0$ $p_i = 1$ $i = 1, 2$	$\frac{b_1^t b_2^{1-t}}{b_1^t + b_2^{1-t}}$

* Spezialfälle der Gammaverteilung

Literaturverzeichnis

- [1] Adhikari, B.B.; Joshi, D.D.: Distance discrimination et résumé exhaustif, *Publs. inst. statist. univ. Paris*, Vol. 5, Fasc. 2, S. 57 - 74.
- [2] Bauer, H.: *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, Berlin, de Gruyter, 1968.
- [3] Berger, A.: On uniformly consistent tests, *Ann. Math. Stat.*, Vol. 22, 1951, S. 289 - 293.
- [4] Bhattacharya, A.: *Bull. Cal. Math. Soc.*, 35, 1943, S. 99.
- [5] Bhattacharya, A.: On a measure of divergence between two multinomial populations, *Sankhya*, Bd. 7, 1946, S. 401 - 406.
- [6] Chernoff, H.: A measure of asymptotic efficiency for tests of a hypotheses based on the sum of observations, *Ann. Math. Stat.*, Bd. 23, S. 493 - 507.
- [7] Csiszár, I.: Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen, *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, Bd. 7, 1962, S. 137 - 158.
- [8] Csiszár, I.; Fischer, J.: Informationsentfernungen im Raum der Wahrscheinlichkeitsverteilungen, *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, Bd. 7, 1962, S. 159 - 180.
- [9] Csiszár, I.: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten, *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, Bd. 8, 1963, S. 85 - 108.
- [10] Csiszár, I.: Über topologische und metrische Eigenschaften der relativen Information der Ordnung α , *Trans. of the Third Prague Conference*, Prag 1964, S. 63 - 73.
- [11] Doob, J.L.: *Stochastic Processes*, New York, Wiley, 1964.
- [12] Erwe, F.: *Differential- und Integralrechnung I*, Mannheim, Bibliographisches Institut, 1962.

- [13] Feinstein, A.: Foundations of information theory, New York, Mac-Graw-Hill, 1958.
- [14] Fichtenholz, G.M.: Differential- und Integralrechnung, Bd. I, II, III, VEB Deutscher Verlag der Wissenschaften, 1964.
- [15] Fisz, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik, VEB Deutscher Verlag der Wissenschaften, 1971.
- [16] Ghurye, S.G.: Information and Sufficient subfields, Ann. Math. Stat., Bd. 38, 1968, S. 2056 - 2065.
- [17] Golomb, S.W.: The information generating function of a probability distribution, Trans. Fourth Prague Conference 1965, S. 291 - 295.
- [18] Grell, H. (Herausgeber): Arbeiten zur Informationstheorie, Bd. I - V, Deutscher Verlag der Wissenschaften, Berlin, 1957.
- [19] Halmos, P.R.; Savage, L.J.: Application of the Radon-Nikodým theorem to the theory of sufficient statistics, Ann. Math. Stat., Bd. 20, 1949, S. 225 - 241.
- [20] Halmos, P.R.: Measure Theory, New York, van Nostrand, 1950.
- [21] Hannan, J.: Consistency of maximum-likelihood estimation of discrete distributions, Essays in Honor of H. Hotelling, Stanford Univ. Press, 1960, S. 249 - 257.
- [22] Hardy, G.H.; Littlewood, J.E.; Pólya, G.: Inequalities, Cambridge at the University Press, 1967.
- [23] Hewitt, E.; Stromberg, K.: Real and abstract analysis, Berlin, Heidelberg, New York, Springer 1965.
- [24] Hoeffding, W.; Wolfowitz, J.: Distinguishability of sets of distributions, Ann. Math. Stat., Bd. 29, 1958, S. 249.
- [25] Hogg, R.V.; Craig, A.T.: Introduction to mathematical statistics, New York 1965, The Macmillan Company.
- [26] Ikeda, S.: Continuity and characterization of Shannon-Wiener information measure for continuous probability distributions, Ann. Inst. Stat. Math., Bd. 11, 1959, S. 131 - 144.

- [27] Ikeda, S.: A remark on the convergence of Kullback-Leibler's mean information, Ann. Inst. Stat. Math., Bd. 12, 1960, S. 81 - 88.
- [28] Ikeda, S.: An application of the discrimination information measure to the theory of testing hypotheses, Part I: Ann. Inst. Stat. Math., Bd. 13, 1961, S. 27 - 46, Part II: Ann. Inst. Stat. Math., Bd. 13, 1961, S. 61 - 89.
- [29] Ikeda, S.: A note on the characterization of Shannon-Wiener's measure of information, Ann. Inst. Stat. Math., Bd. 13, 1962, S. 259 - 266.
- [30] Ikeda, S.: On characterization of the Kullback-Leibler-mean-information for continuous probability distributions, Ann. Inst. Stat. Math., Bd. 14, 1962/63, S. 73 - 79.
- [31] Ikeda, S.: Necessary conditions for the convergence of Kullback-Leibler's mean information, Ann. Inst. Stat. Math., Bd. 14, 1962/63, S. 107 - 118.
- [32] Ikeda, S.: Asymptotic equivalence of probability distributions with applications to some problems of asymptotic independence, Ann. Inst. Stat. Math., Bd. 15, 1963, S. 87 - 116.
- [33] Ikeda, S.: A remark on the incomparability of two criteria for a uniform convergence of probability measures, Ann. Inst. Stat. Math., Bd. 21, 1969, S. 533 - 536.
- [34] Kakutani, S.: On equivalence of infinite product measures, Annals of Mathematics, Bd. 49, 1948, S. 214 - 222.
- [35] Kallianpur, G.: On the amount of information contained in a σ -field, Essays in Honor of H. Hotelling, Stanford University Press, 1960, S. 265 - 271.
- [36] Kingman, J.F.C.; Taylor, S.J.: Introduction to measure and probability, Cambridge at he University Press, 1966.
- [37] Knopp, K.: Theorie und Anwendung der unendlichen Reihen, Springer-Verlag Berlin, 1964.
- [38] Kraft, C.: Some conditions for consistency and uniform consistency of statistical procedures, Univ. Calif. Publ. Stat., 1955, Bd. 2, S. 125 - 142.

- [39] Krickeberg, K.: Wahrscheinlichkeitstheorie, B. G. Teubner Verlagsgesellschaft, Stuttgart, 1963.
- [40] Kullback, S.: Information theory and statistics, New, York, Wiley, 1959.
- [41] Lindley, D.V.: On a measure of the information provided by an experiment, Ann. Math. Stat., Bd. 27, 1956, S. 986 - 1005.
- [42] Matusita, K.: On the theory of statistical decision functions, Ann. Inst. Math. Stat., Bd. 3, 1951, S. 17 - 35.
- [43] Matusita, K.; Akaike, H.: Note on the decision problem, Ann. Inst. Math. Stat., Bd. 4, 1952, S. 11 - 14.
- [44] Matusita, K.: Decision rule, based on the distance, for the classification problem, Ann. Inst. Stat. Math., Bd. 6, 1954/55, S. 613 - 640.
- [45] Matusita, K.; Suzuki, Y.; Hudimoto, H.: On testing statistical hypotheses, Ann. Inst. Stat. Math., Bd. 6, 1954/55, S. 133 - 141.
- [46] Rényi, A.: On measures of entropy and information, Proc. 4th Berk. Symp., Bd. I, 1961, S. 547 - 561.
- [47] Rényi, A.: On the foundations of information theory, Review Int. Stat. Inst., Bd. 33, 1965, 1.
- [48] Rényi, A.: Wahrscheinlichkeitsrechnung, Berlin, 1966, VEB Deutscher Verlag der Wissenschaften.
- [49] Särndal, C.E.: A class of explicata for "information" and a weight of evidence, Review Int. Inst. Stat., Bd. 38, 2, 1970.
- [50] Särndal, C.E.: Properties of some simple functions of the likelihood ratio, Austral. J. Statist. 14 (1), 1972, S. 41 - 49.
- [51] Schmetterer, L.: Mathematische Statistik, Springer-Verlag Wien - New York, 1966.
- [52] Skarabis, H.: Mathematische Grundlagen und praktische Aspekte der Diskrimination und Klassifikation, Physica-Verlag, Würzburg 1970.

- [53] Smith, W.L.: A note on truncation and sufficient statistics,
Ann. Math. Stat., Bd. 28, 1957, S. 247 - 252.
- [54] Wetzel, W.; Jöhnk, M.D.; Naeve, P.: Statistische Tabellen,
Berlin, 1967, Walter de Gruyter.
- [55] Wilks, S.S.: Mathematical statistics, New York, Wiley, 1972.
- [56] Witting, H.: Mathematische Statistik, Stuttgart, Teubner, 1969.
- [57] McShane, E.J.: Integration, Princeton University Press, 1944.

Lebenslauf

- Geboren: Am 14.7.1943 als Sohn des Unteroffiziers Gotthard Trenkler und dessen Ehefrau Marianne, geb. Frauendorf, in Dresden.
- Schule: Besuch der Grundschule in Taucha bei Leipzig von 1949 - 1957.
Flucht aus politischen Gründen 1957.
Besuch des Eckener-Gymnasiums in Berlin-Mariendorf von 1957 - 1962.
- Abitur: September 1962.
- Studium: Chemie 1962 - 1965 Freie Universität Berlin.
Abbruch aus gesundheitlichen Gründen.
Mathematik 1965 - 1970 Freie Universität Berlin.
Vordiplom 1968.
Diplom 1970.
- Wiss. Tutor: 1968 - 1970 am Fachbereich 10 der Freien Universität Berlin für mathematische Propädeutik.
- Wiss. Assistent: Seit 1970 am Fachbereich 10 der Freien Universität Berlin beim Institut für Angewandte Statistik.

